

VOLUME 103

PART C NUMBER 4

SEPTEMBER 1956



The Proceedings
OF
THE INSTITUTION OF
ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C
MONOGRAPHS Nos. 159-186

SAVOY PLACE • LONDON W.C.2

Price Eight Shillings and Sixpence

The Institution of Electrical Engineers

FOUNDED 1871

INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1955-1956

President

SIR GEORGE H. NELSON, BART.

Past-Presidents

SIR JAMES SWINBURNE, BART., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
P. V. HUNTER, C.B.E.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).

COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C.,
T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
J. ECCLES, C.B.E., B.Sc.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.

SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D.,
B.Sc.
SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH.

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
C. M. COCK.
A. R. COOPER, M.Eng.
A. T. CRAWFORD, B.Sc.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
F. J. LANE, O.B.E., M.Sc.
G. S. C. LUCAS, O.B.E.
D. McDONALD, B.Sc.

C. T. MELLING, C.B.E., M.Sc.Tech.
H. H. MULLENS, B.Sc.
W. F. PARKER.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D.
G. L. WATES, J.P.
G. O. WATSON.
D. B. WELBOURN, M.A.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.
R. T. B. WYNN, C.B.E., M.A.

Chairman and Past-Chairmen of Sections

Measurement and Control:

W. BAMFORD, B.Sc.
*M. WHITEHEAD.

Radio and Telecommunication:

H. STANESBY.
*C. W. OATLEY, O.B.E., M.A., M.Sc.

Supply:

L. DRUCQUER.
*J. D. PEATTIE, C.B.E., B.Sc.

Utilization:

D. B. HOGG, M.B.E., T.D.
*J. I. BERNARD, B.Sc.Tech.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:

F. R. C. ROBERTS.
*J. H. MITCHELL, B.Sc., Ph.D.

Mersey and North Wales Centre:

PROFESSOR J. M. MEEK, D.Eng.
*P. R. DUNN, B.Sc.

North Midland Centre:

F. BARRELL.
*G. CATON.

North-Eastern Centre:

A. H. KENYON.
*G. W. B. MITCHELL, B.A.

North-Western Centre:

G. V. SADLER.
*PROFESSOR E. BRADSHAW, M.B.E.,
M.Sc.Tech., Ph.D.

Northern Ireland Centre:

MAJOR E. N. CUNLIFFE, B.Sc.Tech.
*MAJOR P. L. BARKER, B.Sc.

Western Centre:

T. G. DASH, J.P.
*A. N. IRENS.
* Past-Chairman.

Scottish Centre:

E. WILKINSON, Ph.D., B.Eng.
*J. S. HASTIE, B.Sc.(Eng.).

South Midland Centre:

H. S. DAVIDSON, T.D.
*A. R. BLANDFORD.

Southern Centre:

L. H. FULLER, B.Sc.(Eng.).
*E. A. LOGAN, M.Sc.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers. An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 103. PART C. No. 4.

SEPTEMBER 1956

DISCUSSION ON

'THE SUPPRESSION OF SWITCHING TRANSIENTS BY A SHUNT RC CIRCUIT'*

Dr. J. C. Read (*communicated*): One of the possible applications of the shunt RC damping circuit discussed in the monograph is for reducing inverse voltage overswing in extra-high-voltage rectifiers, which results from commutation oscillations (and from hole storage effects in semi-conductor rectifiers, such as germanium rectifiers). It may therefore be of interest to recall, for the sake of completeness, that the main characteristics of this shunt RC circuit were published in a paper of mine in 1945,[†]

in eqns. (8)–(12). In that paper the damping circuits were only a minor incident in a larger subject, so for brevity the derivation of the formulae was not given. The author's monograph has performed a useful purpose in remedying this omission.

Mr. E. A. Finlay (*in reply*): I wish to thank Dr. Read for drawing my attention to his paper, in which the conditions derived in the monograph for the occurrence of non-oscillatory transients had already been stated. I regret that no reference was made in the monograph to his paper, which was not known to me at the time the monograph was written.

*FINLAY, E. A.: Monograph No. 97 U, May, 1954 (see 101, Part IV, p. 266).
†READ, J. C.: 'High-Voltage Steel-Tank Mercury-Arc Rectifier Equipments for Radio Transmitters', *Journal I.E.E.*, 1945, 92, Part II, p. 453.

DISCUSSION ON

'MATRIX METHODS FOR THE EVALUATION OF SIMULTANEOUS FAULTS IN THREE-PHASE SYSTEMS'*

Mr. S. R. Deards (*communicated*): I should be interested to know how the well-known rule relating to the physical significance of the components of a network loop-impedance matrix comes to be attributed to Mr. Austen Stigant. The rule dates back to Maxwell^A and has since appeared in one form or another in the writings of Campbell,^B Carson,^C Guillemin,^D Burington^E and Bode,^F and can be found, together with the dual rule for the node-admittance matrix, in almost every book on electrical network analysis.

(A) MAXWELL, J. C.: 'A Treatise on Electricity and Magnetism' (Oxford, 1892), Vol. I, 3rd edition, Chapter VI, Art. 282b.

(B) CAMPBELL, G. A.: 'Cisoidal Oscillations', *Transactions of the American I.E.E.*, 1911, 30, p. 873.

* LEWIS, W. E., and BANKS, J. H.: Monograph No. 125 S, April, 1955 (see 102 C, p. 231).

(C) CARSON, J. R.: 'Electric Circuit Theory and the Operational Calculus' (McGraw-Hill Book Co., Inc., New York, 1926), Chapter I.

(D) GUILLEMIN, E. A.: 'Communication Networks' (John Wiley and Sons, Inc., 1931), Vol. I, Chapter IV, Art. 3.

(E) BURINGTON, R. S.: 'Matrices in Electric Circuit Theory', *Journal of Mathematics and Physics*, 1935, 14, No. 4, p. 325.

(F) BODE, H. W.: 'Network Analysis and Feedback Amplifier Design' (D. van Nostrand Co., Inc., New York, 1945), Chapter I.

Messrs. W. E. Lewis and J. H. Banks (*in reply*): As this rule is of great use in matrix analysis it would be as well if further suggestions were made in order that an agreed name or title for it could be established.

RELATIVE DYNAMIC STABILITY OF LARGE SYNCHRONOUS GENERATORS

By H. K. MESSERLE, M.Eng.Sc., B.E.E.

(The paper was received 16th June, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

The steady-state stability limit of a synchronous machine can be modified considerably by controlling field excitation and prime-mover torque by means of regulators and governors. In the paper the effects of the individual alternator parameters and controller constants on the dynamic stability limit (i.e. the steady-state limit of a controlled alternator) are investigated, and optimum operating conditions for various types of regulator and governor are established.

The investigation covers a general comparison of the relative dynamic stability of the common types of large alternator. In addition, the variation in dynamic stability with a change in output rating is discussed.

LIST OF SYMBOLS

(a) For Alternator.

$p = \frac{d}{dt}$ = Time derivative.

ΔP_m = Variation in prime-mover output.

ΔP_{el} = Variation in alternator output.

$\Delta \delta$ = Variation in direct axis of alternator with respect to stator.

$M = 4\pi f_0 H$.

H = Per-unit inertia constant.

f_0 = Frequency (50 c/s throughout).

D = System damping plus effect of damper windings.

v = Applied field voltage.

v_{fd} = Equivalent field voltage = $X_{afd}v/R_f$.

i_f = Field current; $i_{fd} = X_{afd}i_f$.

Φ_{fd} = Field flux; $\Phi_{fd} = X_{afd}\Phi_{ff}/X_f$.

X_f = Reactance of field winding.

X_{afd} = Mutual reactance between field and direct-axis armature windings.

$\tau'_{do} = X'_f/R_f$ = open-circuit alternator field time-constant.

R_f = Field-winding resistance.

X_d = Direct-axis synchronous reactance.

X_q = Quadrature-axis synchronous reactance.

X'_d = Direct-axis transient reactance.

Z_e = External impedance.

$\tau'_{dz} = \frac{X'_d + X_e}{X_d + X_e} \tau'_{do}$ = Effective field time-constant.

τ_m = Period of natural mechanical vibrations of turbo-alternator set as determined by inertia and electrical restoring torque.

v_t = Terminal voltage.

v_2 = Infinite busbar voltage.

v'_q = Voltage behind transient reactance.

v_d = Direct-axis component of v_r .

v_q = Quadrature-axis component of v_r .

i = Armature current.

d = Direct-axis component of i .

i_q = Quadrature-axis component of i .

(b) For Regulator.

τ_e = Exciter field time-constant.

τ_s = Stabilizer time-constant.

μ_s = Stabilizer ratio (effective).

μ_{so} = Optimum stabilizer ratio = $\mu_{smin} + k_s$.

μ_{smin} = Minimum value of μ_s for which a regulator provides stable operation.

k_s = Constant.

μ' = Amplifier gain.

μ = Effective overall feedback gain.

μ_o = Optimum feedback gain = $\mu_{min_o} + \frac{1}{3}(\mu_{max_o} - \mu_{min_o})$.

μ_{max_o} = Maximum stable μ for $\mu_s = \mu_{so}$.

μ_{min_o} = Minimum stable μ for $\mu_s = \mu_{so}$.

τ_a = Amplifier time-constant.

τ_c = Compensator time-constant.

μ_c = Compensator ratio.

s_o = Optimum stability ratio = $\frac{\mu_{max_o}}{\mu_{min_o}}$.

S_o = Optimum operating condition in $\langle \mu, \mu_s \rangle$ plane.

(c) For Governor.

r = Regulation.

F_c = Subsidiary controller transfer function.

α_1, α_2 = Subsidiary controller constants.

τ_1, τ_2 = Governor time-constants.

τ_3 = Subsidiary controller time-constant.

$\Delta \omega$ = Variation in alternator speed with respect to synchronous speed.

(1) INTRODUCTION

The steady-state stability of synchronous alternators can be modified by controlling prime-mover torque and field voltage, and the steady-state stability limit of a controlled alternator defined as the dynamic stability limit. This limit is determined largely by the feedback parameters introduced by governors and continuously acting regulators. Very important also are the alternator characteristics, the load and the system to which the alternator is connected.

In recent years particular attention has been focused on the dynamic limit for leading-power-factor operation following trends in the developments in modern power systems.^{1,2} The effects of several types of load system on the dynamic limit of an alternator have been investigated in some recent publications^{1,2} considering the leading-power-factor region. In particular, the improvement that can be achieved by using field continuous output-voltage regulation has been of special interest. The effect of some of the feedback parameters introduced by regulators has been discussed for a few typical machines operating at unity power factor,^{3,4} and methods for investigating the combined effects of governors controlling the prime-mover torque and regulators controlling the alternator field excitation have been investigated in a separate paper.⁴

In the present paper the differences in dynamic stability of the various types of available alternators is discussed in relation to the differences in their construction and output rating. The

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Messerle is at the University of Sydney, New South Wales.

analysis is based on the methods mentioned,⁴ using the general theory for synchronous machines and allowing for the effects of governors and regulators.

The effects of changing feedback and alternator parameters on the dynamic stability limit will be investigated, using as a first example a typical 30 MW turbo-alternator. For any given type of alternator, however, only small differences in the main parameters arise in practice,^{5,6} even with alternators from various manufacturers. Thus any big variation in the parameters on a per-unit basis means effectively a change in output rating or construction, and so the analysis of the 30 MW alternator leads to a comparison of the relative dynamic stability of various types of alternator for the same and for different output ratings.

The average values of two important alternator parameters are plotted in Figs. 3A and 3B, which show the change in alternator-field time-constant τ'_{do} and inertia constant H with output rating. These Figures are based on data obtained from References 5 and 6 and from various manufacturers. In general, the most important parameters are:

- (a) The open-circuit alternator-field time-constant τ'_{do} and the effective field time-constant τ'_{dz} , related by the expression

$$\tau'_{dz} = \frac{X'_d + X_e \tau'_{do}}{X_d + X_e}$$

where X_e is the external reactance between the alternator and the main system if the alternator is assumed to feed into an infinite busbar through the reactance X_e .

- (b) The per-unit inertia constant H , which is defined as

$$H = \frac{\frac{1}{2} \mathcal{J} \omega_0'^2}{G} = \frac{\text{stored energy (joules)}}{\text{rating (volt-amperes)}}$$

where \mathcal{J} = inertia of alternator set including prime mover, allowance being made for rotating water inertia in water-wheel drives.

$\omega_0' = 2\pi f_0'$, the rated speed in revolutions per second.
 G = rated output of alternator.

- (c) Alternator reactances.

Of these parameters, τ'_{do} and H will be investigated in detail in the paper; in general it will be shown that an increase in τ'_{do} or a reduction in H improves the stability of an alternator. The effect of alternator reactances has been discussed in Reference 1, and it is shown that the stable operating range for an alternator can be extended by reducing any of the alternator reactances. However, in practice a small reactance requires a large frame size which makes the design very expensive, particularly for the higher ratings. The high reactances of very large alternators causing unstable operation in modern systems are often the main reason for the use of fast and continuously acting automatic voltage-regulators.

For the calculations presented in the paper the author has had the use of two automatic computers: (a) the mechanical differential analyser of the Mathematical Instruments Section, C.S.I.R.O., Sydney, and (b) the electronic aircraft simulator of the Aeronautical Engineering Department, University of Sydney.

(2) ALTERNATOR ANALYSIS

A very important consideration in this investigation is the mathematical representation of the alternator. The representation here is based on the analysis given in Reference 4, and the variations in the alternator currents and voltages are related as shown in Fig. 1. This is the vector diagram for a synchronous alternator, on the assumption that variations in alternator speed during small disturbances have a negligible effect on the electrical

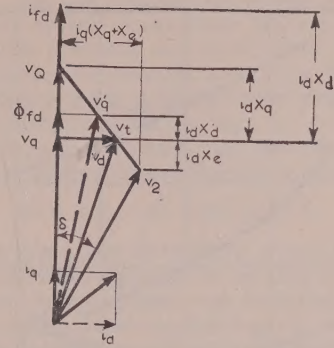


Fig. 1.—Vector diagram relating alternator voltages and currents.

variables. Dumper windings are allowed for in the power equation, which is

$$Mp^2 \Delta \delta + Dp \Delta \delta = \Delta P_m - \Delta P_{el} \quad (1)$$

The effects of transients and transient saliency are allowed for by considering the relation between field current Δi_f and applied field voltage Δv , where the symbol Δ always implies small variations in the variable concerned. Thus

$$\Delta v = R_f \Delta i_f + p \Delta \Phi_{fd} \quad (2)$$

When per-unit notation is used eqn. (2) can be changed into

$$\Delta v_{fd} = \Delta i_{fd} + \tau'_{do} p \Delta \Phi_{fd} \quad (3)$$

where

$$\Delta v_{fd} = \tau'_{do} \Delta v$$

$$\Delta i_{fd} = X_{afd} \Delta i_f$$

$$\Delta \Phi_{fd} = \frac{X_{afd}}{X_f} \Delta \Phi_{ff}$$

The magnitude of Φ_{fd} is related to the other variables as shown in Fig. 1. A detailed derivation of these relations is given in Reference 4.

Any control quantities from the regulator and governor can be introduced in eqns. (1) and (3). The mechanical input torque is controlled by the governor, and the control effect can be allowed for in eqn. (1) as a variation ΔP_m . A regulator controls the field voltage, which is Δv_{fd} in eqn. (3).

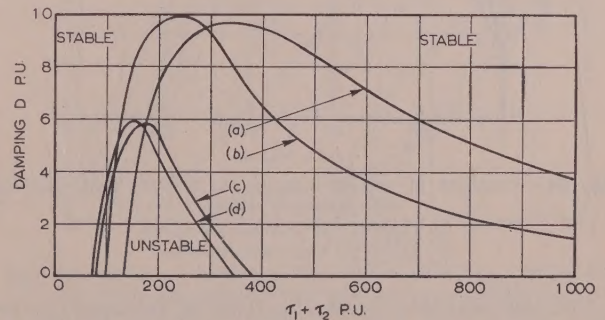


Fig. 2.—Damping required for stability of a typical 30 MW alternator with speed governor under following assumed conditions.

- (a) $X_d = X_q = X'_d = 1.39$ p.u. at 1.0 p.f. and 1.0 p.u. current.
 (b) $X_d = X_q = X'_d = 0.22$ p.u. at 1.0 p.f. and 1.0 p.u. current.
 (c) $X_d = X_q = 1.39$, $X'_d = 0.22$ p.u. at 1.0 p.f. and 1.0 p.u. current.
 (d) $X_d = X_q = 1.39$, $X'_d = 0.22$ p.u. at 1.0 p.f. and 1.3 p.u. current.
 Regulation $r = 1.25\%$ in all cases.

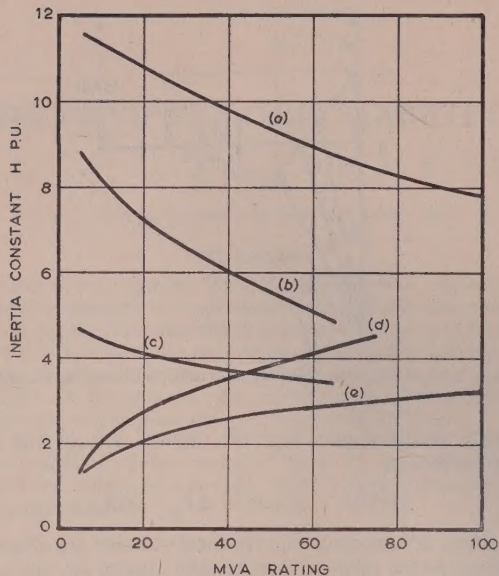


Fig. 3A.—Variation of average inertia constant H with generator rating.

- (a) 1500 r.p.m. } turbo-alternators, condensing.
 (b) 3000 r.p.m. }
 (c) 3000 r.p.m. } turbo-alternators, non-condensing.
 (d) 500 r.p.m. } water-wheel generator allowing 15% for water wheel.
 (e) 100 r.p.m. }

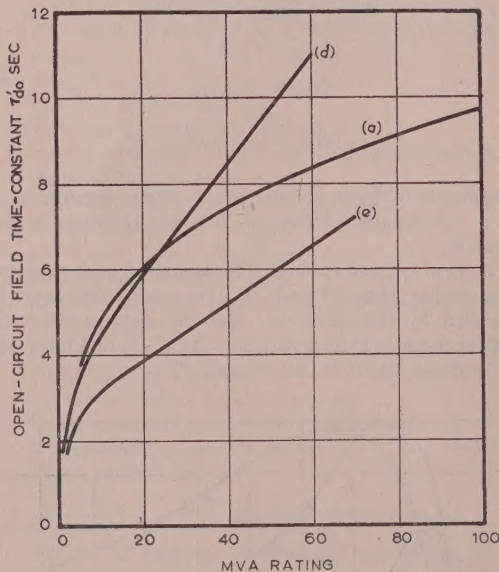


Fig. 3B.—Variation of average open-circuit field time-constant τ'_{do} with generator rating.

- (a) 1500 and 3000 r.p.m. turbo-alternators.
 (b) 500 r.p.m. } water-wheel generators.
 (c) 100 r.p.m. }

The usual practice is to use a much simpler representation for the alternator. However, as will be seen from Fig. 2, any simplification in the machine analysis has to be considered with care. The curves in Fig. 2 provide the damping D required [see eqn. (1)] to stabilize a typical 30 MW alternator which is controlled by a speed governor,⁸ and D is plotted against $(\tau_1 + \tau_2)$, the overall time delay in the governor; curve (c) has been computed by the use of the generalized approach given

above for the alternator operating at unity power factor and 1.0 p.u. current.

Neglecting transient reactance and representing the machine by its synchronous reactance, the analysis becomes much simpler. This, however, leads to a considerable difference in the value of D required for stabilization, as follows from curve (a) in Fig. 2 when compared with curve (c). Curve (a) was computed by increasing X'_d and making it equal to X_d . It follows that the general steady-state representation of an alternator can provide only very rough results.

To allow for transient properties of the alternator the magnitudes of X'_d and X'_q are usually reduced and made equal to the transient reactance X'_d . This assumption is used for curve (b) (Fig. 2), and again there is a very marked difference when comparing this curve with curve (c). Thus it was found necessary to use the generalized approach for the work discussed in this paper.

Another important factor is the load system into which the alternator operates. In particular, if various alternators are to be compared they should be operating under similar conditions. Usually, most of the larger alternators are connected to a large distribution system which can be approximated by an infinite busbar. The alternator can then be assumed to be connected to an infinite busbar through an external impedance Z_e . An local load or intermediate load close to the alternator terminal can be allowed for in Z_e by the use of Thevenin's theorem. The value of Z_e is usually mainly inductive, and varies between 0.2 and 0.8 p.u., according to the rating of the alternator. For the analysis here, an average value was taken, such that

$$Z_e = jX_e = j0.4 \text{ p.u.}$$

The characteristic constants of the 30 MW turbo-alternator (condensing type) used as a basis of this investigation are (unless stated otherwise)

$$\begin{aligned} H &= 6.0 & \tau'_{dz} &= 650 \\ D &= 3.0 & X_d &= X_q = 1.39 \\ \tau'_{do} &= 1880 \text{ p.u.} = 5.8 \text{ sec} & X'_d &= 0.22 \\ v_t &= 1.0 \end{aligned}$$

All quantities are in per-unit notation based on rated power and voltage of the alternator.

(3) EFFECT OF ALTERNATOR AND CONTROL PARAMETERS ON DYNAMIC STABILITY LIMIT

(3.1) Constant Prime-Mover Torque

(3.1.1) Variation of Regulator Constants.

A fast continuously acting regulator controlling field excitation usually improves the dynamic stability of the alternator but the actual performance depends largely upon the speed of response. The controlling quantity is usually the output voltage,^{3,4} and in that case the main purpose of a regulator is to keep the output voltage constant, requiring a high overall feedback gain. The improvement in the steady-state stability limit may be of secondary importance for alternators below about 30 MW. For bigger alternators having large reactances and operating near unity power factor, the steady-state stability limit can become very critical, and regulators may be necessary for stable operation.

A typical regulator is shown in Fig. 4. It generally consists of four components,^{3,4} namely the exciter, the amplifier, the stabilizer, and the error compensator. The amplifier usually consists of a fast-acting amplidyne, metadyne or magnetic amplifier, and thus in most cases the time delay in the amplifier stage can be neglected when compared with other circuit components.⁴

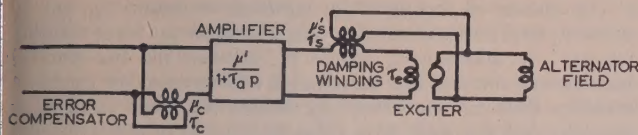


Fig. 4.—A typical regulator with the transfer function

$$F_R = \left(\frac{-\mu}{1 + \tau_a p} \right) \left(1 + \frac{\mu_c \tau_c p}{1 + \tau_c p} \right) \left\{ \frac{1 + \tau_s p}{1 + [\tau_s + \tau_s'(1 + \mu_s)]p + \tau_s \tau_s' p^2} \right\}$$

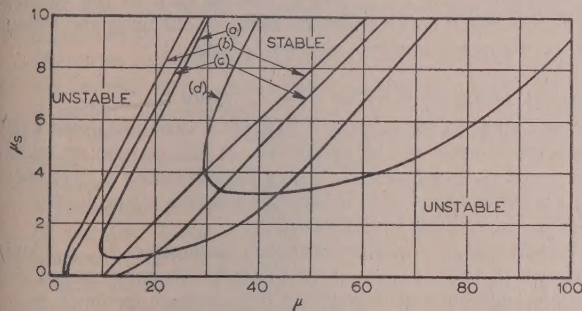
The regulator in Fig. 4 is used here to show the effect of the feedback elements on the dynamic limit. By varying the time-constants and gains it is possible to determine the effect of different regulators. The parameters used for the regulator, unless stated otherwise, are

$$\tau_e = \tau_s = 500 \text{ p.u. (approximately } 1.6 \text{ sec)}$$

$$\mu_c = \tau_c = 0$$

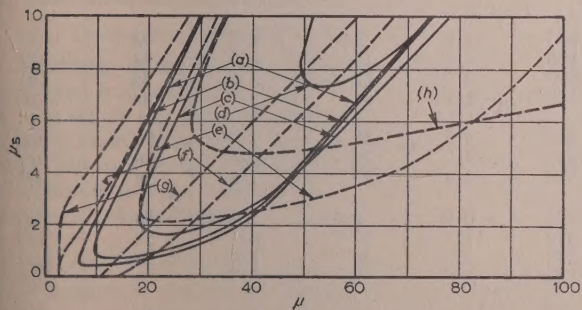
$$\tau_a = 30 \text{ p.u. (approximately } 0.1 \text{ sec)}$$

As shown in Figs. 5A–5D the effective feedback gains μ and the stabilizer constants μ_s for which the alternator is stable can now


 Fig. 5A.—Effect of exciter time-constant τ_e on stability ranges for voltage-regulated alternator, at unity power factor and 2.2 p.u. current.

$$H = 6; D = 3.0; v_{t0} = 1.0; X_d = X_q = 1.39; X'_d = 0.22; X_e = 0.4; \tau'_{d0} = 1880; \tau_s = 500.$$

- | | |
|---|------|
| (a) $\tau_e = 500 \text{ p.u. (= } 1.6 \text{ sec).}$ | p.u. |
| (b) $\tau_e = 125$ | |
| (c) $\tau_e = 250$ | |
| (d) $\tau_e = 1000$ | |


 Fig. 5B.—Effect of exciter time-constant τ_e and stabilizer time-constant τ_s on stability ranges for voltage regulated alternator, at unity power factor and 2.2 p.u. current.

$$H = 6; D = 3.0; v_{t0} = 1.0; X_d = X_q = 1.39; X'_d = 0.22; X_e = 0.4; \tau'_{d0} = 1880.$$

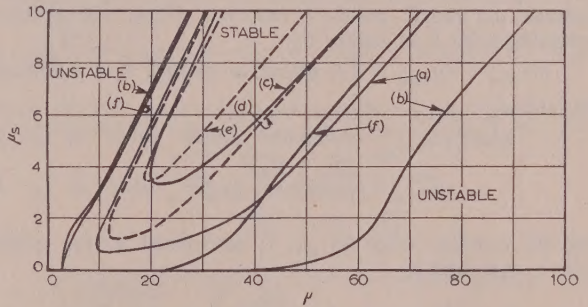
- | | |
|-----------------------------------|------|
| (a) $\tau_e = 500; \tau_s = 500$ | p.u. |
| (b) $\tau_e = 500; \tau_s = 1000$ | |
| (c) $\tau_e = 500; \tau_s = 250$ | |
| (d) $\tau_e = 500; \tau_s = 125$ | |
| (e) $\tau_e = \tau_s = 1000$ | |
| (f) $\tau_e = \tau_s = 250$ | |
| (g) $\tau_e = \tau_s = 125$ | |
| (h) $\tau_e = \tau_s = 2000$ | |

be determined. The load condition considered in Figs. 5A–5D is a 2.2 p.u. current at unity power factor, and the stability curves provide the stable operating regions for different values of μ and μ_s . Considering, for example, curve (a) in Fig. 5A, we find that the operation is stable inside the curve and unstable outside—i.e. for a particular value of μ_s (say, $\mu_s = 4$) the alternator must be unstable for low feedback gains and also for high feedback gains, leaving only a limited stability range for μ , which is

$$16.0 < \mu < 47.8 \text{ (for } \mu_s = 4)$$

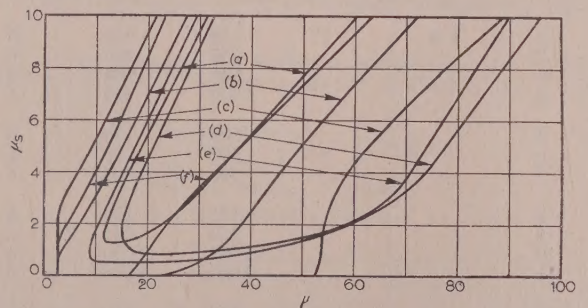
The feedback gain μ and the stabilizer constant μ_s are chosen as the variable parameters in Figs. 5A–5D since μ and μ_s are the only parameters in the system, not excluding those of the alternator, that can be readily changed even on a completed unit. In addition, the diagrams provide a means of defining optimum operating conditions such that the designer of regulators, or the power-system engineer, can determine the relative performance and suitability of different types of regulators:

The general design requirements for a regulator are a high value for μ and a low value for μ_s . In practice the overall feedback gain may become as large as 100, whereas μ_s is usually


 Fig. 5C.—Effect of alternator-field time-constant τ'_{d0} and inertia constant H on stability ranges for voltage-regulated alternator, at unity power factor and 2.2 p.u. current.

$$D = 3.0; v_{t0} = 1.0; X_d = X_q = 1.39; X'_d = 0.22; X_e = 0.4; \tau_e = \tau_s = 500.$$

- | | | |
|-------------------------|------|-----------|
| (a) $\tau'_{d0} = 1880$ | p.u. | $H = 6.$ |
| (b) $\tau'_{d0} = 3760$ | | $H = 6.$ |
| (c) $\tau'_{d0} = 1040$ | | $H = 6.$ |
| (d) $\tau'_{d0} = 1880$ | | $H = 12.$ |
| (e) $\tau'_{d0} = 1880$ | | $H = 24.$ |
| (f) $\tau'_{d0} = 3760$ | | $H = 12.$ |


 Fig. 5D.—Effect of alternator-field time-constant τ'_{d0} , exciter time-constant τ_e and stabilizer time-constant τ_s on stability ranges for voltage-regulated alternator, at unity power factor and 2.2 p.u. current.

$$H = 12; D = 3.0; v_{t0} = 1.0; X_d = X_q = 1.39; X'_d = 0.22; X_e = 0.4$$

- | | |
|--|------|
| (a) $\tau'_{d0} = 1880; \tau_e = \tau_s = 500$ | p.u. |
| (b) $\tau'_{d0} = 3760; \tau_e = \tau_s = 500$ | |
| (c) $\tau'_{d0} = 7520; \tau_e = \tau_s = 500$ | |
| (d) $\tau'_{d0} = 3760; \tau_e = 1000; \tau_s = 500$ | |
| (e) $\tau'_{d0} = 3760; \tau_e = \tau_s = 1000$ | |
| (f) $\tau'_{d0} = 3760; \tau_e = 250; \tau_s = 500$ | |

kept below 10. Thus for optimum performance the stability curves should cover high values of the feedback gain μ at low values of stabilizer constant μ_s , and the stability ranges therefore become a measure of the degree of stability for any combination of machine and regulator parameters.

All stability curves in Figs. 5A and 5B apply to changes in regulators, and the relative dynamic stability can be determined by comparison. To put this comparison into a quantitative form it is necessary to define the optimum operating condition S_o for a regulator.

(3.1.1.1) Optimum Operating Conditions.

The optimum operating point S_o is determined by the shape and location of the stability curve and is defined by the co-ordinates μ_o and μ_{so} . The shapes of all stability curves are very similar; they all have a minimum value μ_{smin} and the high-gain and low-gain branches are nearly parallel in most cases, the slopes being within the range

$$0.2 < \frac{d\mu_s}{d\mu} < 0.4$$

The main difference in the curves lies in μ_{smin} and the distance between the nearly parallel branches. Hence the optimum operating point S_o is defined by

(a) $\mu_{so} = (\mu_{smin} + k_s)$, where, in general, $k_s = 2$ suffices,

and (b) $\mu_o = \mu_{min_o} + \frac{1}{3}(\mu_{max_o} - \mu_{min_o})$

where μ_{max_o} = (maximum stable value of μ for $\mu_s = \mu_{so}$);

μ_{min_o} = (minimum stable value of μ for $\mu_s = \mu_{so}$);

and the stability range for μ_{so} is determined by the stability ratio s_o expressed by

$$s_o = \frac{\mu_{max_o}}{\mu_{min_o}}$$

The choice of the optimum stabilizer constant μ_{so} has to move S_o well into the stability range to allow for some stability margin. In addition S_o should lie between the two parallel branches of the stability curve since this leads to an optimum stability ratio s . However, μ_{so} should be kept as low as possible, and this leads to a value of 2 for k_s in the expression under (a).

Once μ_{so} has been fixed, the choice of the optimum feedback gain μ_o depends on the performance of the alternator within the stability range. As can be seen in Reference 4, the alternator is more oscillatory towards the upper end of the range, and it has been found that best performance cannot be obtained when using an average value for μ_o , but only when μ_o is taken as being approximately one-third of the stability range from μ_{min_o} .

As an example, consider curve (a) in Fig. 5A for which the optimum operating point at S_{oa} is given by

$$\mu_{so} = 2.7$$

and

$$\mu_o = 22$$

The stability range is measured by the ratio μ_{max_o}/μ_{min_o} , which in this case is $s_o = 3.1$.

The values of μ_o , μ_{so} , and s_o will be used in the following discussion when considering the effect of changing the machine and regulator parameters as shown in Table 1.

(3.1.1.2) Variation in Exciter Time-Constant τ_e .

The change in the stability region for a variation in the exciter time-constant τ_e is shown in Fig. 5A. It follows that an increase in τ_e requires an increase in the stabilizer constant μ_s . However, if μ_s is increased as well, it becomes possible to use higher feedback gains, and this is very desirable.

The change in optimum operating parameters μ_o , μ_{so} and s_o is shown in Table 1 (cases 1, 2, 3 and 4). In consequence, there is a considerable improvement in the optimum feedback gain μ_o with increasing τ_e . However, the value of μ_{so} increases rapidly

Table 1

OPTIMUM OPERATING CONDITIONS

Case number	Power factor	Current	Alternator and regulator parameters				Optimum conditions			Figure number	Curve
			τ_e	τ_s	H	τ'_{do}	μ_{so}	μ_o	s_o		
		p.u.	p.u.	p.u.		p.u.					
1	1.0	2.2	125	500	6	1880	1.0	7.2	4.2	5A	(b)
2			250	500	6	1880	1.7	13	3.6	5A	(c)
3			500	500	6	1880	2.7	22	3.1	5A	(a)
4			1000	500	6	1880	5.1	45	2.5	5A	(d)
5			500	125	6	1880	9.3	58	1.4	5B	(d)
6			500	250	6	1880	3.7	29	2.4	5B	(c)
7			500	1000	6	1880	2.4	20	3.8	5B	(b)
8			125	125	6	1880	0.9	6.8	5.8	5B	(g)
9			250	250	6	1880	1.7	12.7	4.3	5B	(f)
10			1000	1000	6	1880	4.2	36	3.4	5B	(e)
11			2000	2000	6	1880	6.8	52	2.7	5B	(h)
12			500	500	12	1880	3.2	20	2.0	5C	(d)
13			500	500	24	1880	5.5	25	1.4	5C	(e)
14			500	500	6	1080	5.3	27	1.73	5C	(c)
15			500	500	6	3760	2	26	10.7	5C	(b)
16			500	500	12	3760	1.8	16	6.2	5D	(b)
17	0.7	1.7	250	500	6	1880	1.8	7	2.1	6	(b)
18			500	500	6	1880	4.5	15	1.7	6	(a)
19			1000	500	6	1880	8.9	30	1.5	6	(c)
20			500	500	6	3760	2.0	12	3.8	6	(e)

at the same time, and another drawback is a decrease in the stability ratio s_o , which implies a relative decrease in the stability margin.

Thus, with an increase in τ_e goes a slight improvement in the stability limit for high values of feedback gain μ and a marked deterioration for low values. The reduction in stability for low values of μ is due to the drop in speed of response of the regulator with increased time-constant τ_e , which leads to hunting of the alternator. The increase in stability for high feedback gain is due to the fact that instability for the high gains is caused by resonance between feedback oscillations in the feedback loop and the natural mechanical vibrations of the alternator, which are determined by the alternator inertia and the electrical restoring torque.⁴ An increase in τ_e reduces the frequency of the feedback oscillations, and so for a larger value of τ_e a larger value of μ is required to produce instability.

(3.1.1.3) Variation in Stabilizer Time-Constant τ_s .

The effect of changing τ_s is shown in Fig. 5B by curves (d), (c), (a), and (b), and in Table 1 as cases 5, 6, 3, and 7. An increase in τ_s considerably reduces the required stabilization μ_{so} as long as $\tau_s < \tau_e$ and the stability ratio s_o improves as well. The optimum gain drops, but this is not very critical because of the increase in s_o .

It therefore follows that the stability performance improves with an increase in τ_s . An increase in τ_s beyond the magnitude of the exciter time-constant τ_e has little effect on the stability range. Thus, for design purposes, the values of τ_s and τ_e should be made equal (see cases 3 and 7 in Table 1).

The effect of altering τ_s and τ_e together (assuming that $\tau_s = \tau_e$) is similar to that of a change in τ_e alone. However, the overall stability improves if both values are made equal, as follows from cases 8, 9, and 10 in Table 1.

For large values of τ_s and τ_e the required stabilization represented by μ_{so} becomes large, whereas for small values of τ_e and τ_s the optimum feedback gain becomes very small. The optimum condition between these two extremes arises for a magnitude of τ_s and τ_e of the order of 500 to 1000 p.u., and this means that the optimum values of τ_s and τ_e are approximately equal to the effective field time-constant of the alternator, $\tau'_{dz} = 650$. It follows that the time-constants in the regulator should be made approximately equal to τ'_{dz} for optimum conditions.

(3.1.2) Variations in Alternator Parameters

(3.1.2.1) Open-Circuit Alternator Field Time-Constant τ'_{do} .

The change in the optimum parameters due to a variation in τ'_{do} follows from cases 14, 3 and 15 in Table 1, and curves (a), (b), and (c) in Fig. 5c. As is shown, an increase in τ'_{do} considerably improves the stability. For example, if τ'_{do} is doubled from 1880 to 3760 p.u. the stability ratio is effectively trebled, the optimum gain μ_o increases from 22 to 26 and the stabilizer constant μ_{so} drops from 2.7 to 2.0.

(3.1.2.2) Inertia Constant H .

An increase in the inertia constant H has an adverse effect, as follows from cases 3, 12, and 13 in Table 1 and curves (a), (d), and (e) in Fig. 5c. The effect is not so marked as for changes in τ'_{do} , although this difference occurs only because the low gain limit is hardly affected by changes in H . A change in inertia alters only the high gain limit, since it alters the natural frequency of the mechanical vibrations.

When changing H and τ'_{do} together, as shown by curves (a) and (f) in Fig. 5c, the upper stability limit is hardly affected if both are increased in the same proportion. This is to be expected, since an increase in H increases the period of the

mechanical vibrations and an increase in τ'_{do} increases the period of the feedback oscillations in the feedback loop. Thus the changes tend to counterbalance. However, the low-gain part of the stability curve is independent of H , as is demonstrated by curves (f) and (b) in Fig. 5c. For low gains, curve (f) differs from curve (a) and corresponds more to curve (b), and curves (f) and (b) apply to the same value of τ'_{do} .

In Fig. 5D the effect of the variations in the parameters is considered for a machine with twice the original per-unit inertia constant. It can be seen that the conclusions drawn so far apply here too, and they can be taken as quite general.

In Figs. 6 and 7 some additional stability curves are plotted

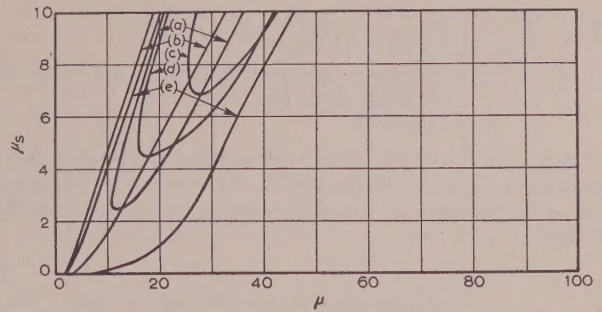


Fig. 6.—Stability-range curves similar to Fig. 5D for 30 MW turbo-alternator at 0.7 power factor and 1.7 p.u. current. ($H = 6$.)

- | | |
|--|--------|
| (a) $\tau'_{do} = 1880$; $\tau_e = \tau_s = 500$ | } p.u. |
| (b) $\tau'_{do} = 1880$; $\tau_e = 250$; $\tau_s = 500$ | |
| (c) $\tau'_{do} = 1880$; $\tau_e = 1000$; $\tau_s = 500$ | |
| (d) $\tau'_{do} = 1880$; $\tau_e = \tau_s = 1000$ | |
| (e) $\tau'_{do} = 3760$; $\tau_e = \tau_s = 500$ | |

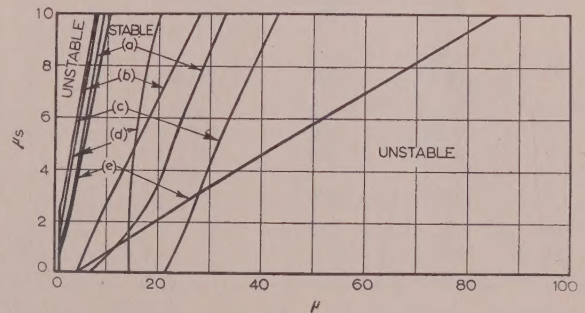


Fig. 7.—Stability ranges for 30 MW turbo-alternator at 0.3 power factor and 1.1 p.u. current as affected by changes in τ_s , τ_e , τ'_{do} , r , τ_1 and τ_2 .

- | | |
|--|-----------------|
| (a) $\tau_s = \tau_e = 500$; $\tau'_{do} = 1880$ p.u. | } no regulator. |
| (b) $\tau_s = 500$; $\tau_e = 125$; $\tau'_{do} = 1880$ p.u. | |
| (c) $\tau_s = \tau_e = 500$; $\tau'_{do} = 3760$ p.u. | |
| (d) $\tau_s = \tau_e = 500$; $\tau'_{do} = 1880$ p.u.; $r = 0.025$; $\tau_1 = \tau_2 = 200$ p.u. | |
| (e) $\tau_s = \tau_e = 500$; $\tau'_{do} = 1880$ p.u.; $r = 0.025$; $\tau_1 = \tau_2 = 0$. | |

for power factors of 0.7 and 0.3, and it is shown that the conclusions mentioned above apply also at different power factors. For lower power factors the stability ranges shift somewhat into a lower range of μ values, as can be seen by considering cases 17 to 20 in Table 1. For constant alternator and regulator parameters, however, the optimum feedback gain, μ_o , does not alter very much even when the power factor is reduced to 0.3.

(3.2) Variable Prime-Mover Torque

(3.2.1) Governors.

The prime-mover torque is, in general, controlled by means of a governor, and the control quantity is usually the alternator

speed. Other control quantities are synchronous time, output power, and tie-line power.

The effect of a governor on the steady-state and dynamic stability limits depends mainly on the governor regulation and overall time delay. So far as the control quantities are concerned, the most effective is the alternator speed, as will be shown.

The simplest mathematical representation of governor performance for a steam turbine involves a second-order time-delay function^{7,8,4} relating the variations in mechanical input torque ΔP_m to the actual control quantities as follows:

$$\Delta P_m = \frac{1}{(1 + p\tau_1)(1 + p\tau_2)} \left(\frac{1}{r} \Delta\omega + F_c \right)$$

$$\text{where } F_c = \frac{1}{(1 + p\tau_3)} \left(\alpha_1 \frac{1}{p} \Delta\omega + \alpha_2 \Delta P_{el} \right),$$

τ_1, τ_2 = Governor time-constants, $(\tau_1 + \tau_2)$ being the overall delay, and

α_1, α_2 = Constants.

The transfer function for a water-turbine governor is more complex, but its main difference lies in the longer overall transient time delay required to prevent excessive shock waves in the penstock.

The general effect of speed governors can be deduced from Fig. 2, which shows the stabilization required for a 30 MW turbo-alternator feeding into an infinite busbar through a reactance $X_e = 0.4$ p.u. Curve (d) is derived for 1.30 p.u. power at unity power factor and shows an unstable operating region for a range of overall time delay in the governor. A maximum amount of damping D is required when $2(\tau_1 + \tau_2)$ is equal to the period of the natural mechanical vibrations of the alternator set, as should be expected. Curve (c) in the same figure has been computed for a lower power output of 1.0 p.u. at unity power factor, and it follows that there is little variation in the unstable region over a wide range of power. The actual steady-state stability limit is at 1.35 p.u. current.

It is interesting to note that the governor has a stabilizing effect, not only for small but also for larger time-constants leaving an unstable region in between. This differs from previously published results, where negative damping is possible only for small time delays.⁸

(3.2.3) Effect of Governor on Dynamic Stability Limit.

The effects of various governor parameters on the dynamic limit of the voltage-regulated alternator are shown in Figs. 7, 8A and 8B. The optimum operating conditions of the

Table 2

OPTIMUM OPERATING CONDITIONS FOR GOVERNOR CONTROL AT 1.0 P.F. AND 2.2 P.U. CURRENT

Case number	Alternator and governor parameters					Optimum conditions			Figure number	Curve
	r	τ_1	τ_2	H	D	μ_{80}	μ_0	s_0		
1	∞	—	—	6	3	2.7	22	3.1	8A	(d)
2	0.025	200	200	6	3	2.0	20	3.8	8A	(b)
3	0.050	200	200	6	3	2.5	21	3.2	8A	(a)
4	∞	—	—	12	3	3.2	20	2.0	8A	(e)
5	0.025	200	200	12	3	2.4	17	2.3	8A	(c)
6	0.05	50	50	6	3	2.8	28	4.5	8B	(d)
7	0.05	100	100	6	3	2.5	24	3.9	8B	(e)
8	0.05	500	500	6	3	2.7	22	3.0	8B	(f)
9	0.05	100	50	6	3	2.6	26	4.1	8B	(g)
10	∞	—	—	6	0	2.7	22	2.9	8B	(a)
11	∞	—	—	6	20	3.0	25	3.3	8B	(c)

Governors can be designed to have time-constants as low as 0.5 sec, and for steam turbines the tendency is to make them as low as possible, since this improves the steady-state stability.⁸ For water turbines the time delay must be at least 5 sec (approximately 1600 p.u.), and this makes water-turbine governors practically ineffective so far as the steady-state and dynamic stability limits of the alternator are concerned.

The regulation r on steam turbines is of the order of 2–15% and, as has been shown,⁸ with a lower value of r , i.e. a smaller speed droop, the alternator operation usually becomes more unstable if there is no regulator controlling the excitation.⁸ For a regulated alternator it is possible to improve the dynamic-stability ranges by reducing r if the governor time delay is low enough, as is shown in Section 3.2.3.1.

(3.2.2) Effect of Governor on Steady-State Stability Limit.

The effect of a speed governor on the steady-state stability of an alternator (i.e. an alternator without a continuously acting regulator) has been investigated in some detail.⁸ However, as is shown in Fig. 2, the simplified approaches that have been used in previous publications should indicate general trends only.

regulator are affected only to a relatively small extent, as can be seen in Table 2, and the limit curves hardly change for low feedback gains even for extreme variations. However, there are changes in the dynamic limit for high feedback gains, since the governor must affect the natural vibrations of the turbo-alternator if its time delay is low enough.

(3.2.3.1) Regulation r .

A variation in the regulation r affects the dynamic limit, shown in Fig. 8A. There the stability curves are computed for normal governor time delay $\tau_1 = \tau_2 = 200$ p.u. (equal to 0.65 sec) and, as may be seen from curves (a), (b), and (d) the changes in the high-gain limits are inversely proportional to the regulation. This conclusion applies to other alternator parameters, as indicated by curves (c) and (e).

(3.2.3.2) Time-Constants τ_1 and τ_2 .

In general, the dynamic stability limit for high-gain feedback increases for the following range of overall time delay:

$$30 \text{ p.u.} < (\tau_1 + \tau_2) < \tau_m$$

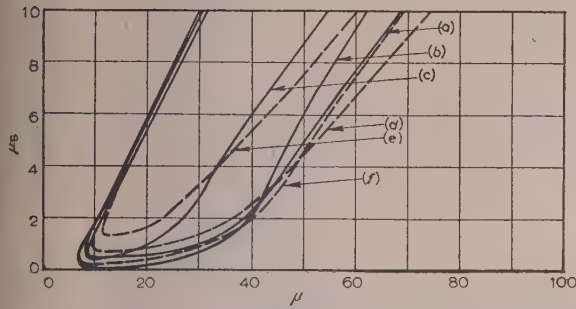


Fig. 8A.—Stability ranges allowing for governor effects, at unity power factor and 2.2 p.u. current, as affected by changes in r and H .

$$\tau'_{do} = 1880; D = 3.0; v_{t0} = 1.0; X_d = X_q = 1.39; X'_d = 0.22; X_e = 0.4; \tau_a = \tau_g = 500.$$

- (a) $r = 0.050; \tau_1 = \tau_2 = 200$ p.u.; $H = 6$.
 (b) $r = 0.025; \tau_1 = \tau_2 = 200$ p.u.; $H = 6$.
 (c) $r = 0.025; \tau_1 = \tau_2 = 200$ p.u.; $H = 12$.
 (d) $r = 0$ (no governor); $H = 6$.
 (e) $r = 0$ (no governor); $H = 12$.
 (f) $r = 0.050; \tau_1 = \tau_2 = 200$ p.u.; $H = 6; \alpha_1 = 0.001$.

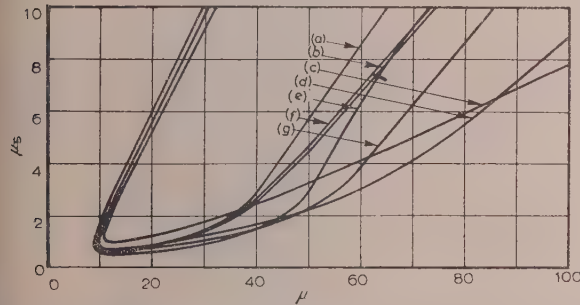


Fig. 8B.—Stability ranges similar to Fig. 8A, as affected by changes in τ_1, τ_2 and D . ($H = 6$)

- (a) $D = 0$
 (b) $D = 3$ } no governor.
 (c) $D = 20$
 (d) $D = 3; r = 0.05; \tau_1 = \tau_2 = 50$
 (e) $D = 3; r = 0.05; \tau_1 = \tau_2 = 100$
 (f) $D = 3; r = 0.05; \tau_1 = \tau_2 = 500$
 (g) $D = 3; r = 0.05; \tau_1 = 100; \tau_2 = 50$ } p.u.

which is the period of natural mechanical vibrations of turbo-alternator set, as appears from Fig. 8B. If the time delay increases beyond the period of the natural mechanical vibrations a reduction in the stability ranges for high stabilizer constants μ_s follows. However, the governor effect becomes negligible as soon as $\tau_1 + \tau_2 > 3\tau_m$ as can be seen in Fig. 8B, curves (b) and (f).

(3.2.3.3) Subsidiary Controller Constants.

Curve (f) in Fig. 8A shows the effect of integrated speed control due to the factor

$$\alpha_1 \frac{1}{p} \Delta \omega = \alpha_1 (\Delta \delta - \Delta \delta_0)$$

where $\Delta \delta$ = Variation in the electrical power angle.

$\Delta \delta_0$ = Initial value.

$\alpha_1 = 0.001$ in this example.

A relatively large value of α_1 is chosen, since, for practical constants (which are of the order of one-tenth of this value⁷), there is hardly any noticeable effect on the dynamic limit.

Despite the large value of α_1 only a small variation in the dynamic limit arises, even when assuming $\tau_3 = 0$, i.e. no time delay in the subsidiary controller.

The effect of electrical power control is much the same as that of integrated speed control, since the electrical torque variation is approximately proportional to $\Delta \delta$, the variation of the machine power angle with respect to the infinite busbar.⁴ It therefore follows that the subsidiary controls under normal conditions have only a very minor effect on the dynamic stability limit.

(4) APPLICATION OF RESULTS TO ALTERNATORS OF NORMAL DESIGN

The discussion so far has been based on a 30 MW 50 c/s condensing type turbo-alternator, with particular reference to the effects of variations in the alternator and feedback parameters. The next step is to apply the conclusions to a comparison of different alternators and to find a basis for a general investigation.

As follows from Section 3, the effects due to the alternator parameters (i.e. τ'_{do} and H) are practically independent of the variation in feedback parameters. This means, for example, that an increase in the alternator-field time-constant τ'_{do} considerably improves the dynamic stability for all practical regulator and governor combinations. Similarly, a reduction in the per-unit inertia H effectively improves the optimum operating conditions for any set of feedback parameters. Thus, knowing the magnitudes of H and τ'_{do} for different alternators we can compare their relative dynamic stability simply by making use of Table 1 and Figs. 3A and 3B.

(4.1) Comparison of Various Types of Alternators of the same Output Rating

(4.1.1) Steam Turbo-Alternators.

Alternators having the same output capacity are compared first. For example, two 30 MW condensing-type turbo-alternators, one with two poles and the other with four poles, have the same field time-constant, as follows from Fig. 3B, but the per-unit inertia constant is 60% higher for the slower machine. Considering now cases 3 and 12 in Table 1 we find a deterioration in the optimum operating conditions for the 4-pole alternator. The optimum feedback gain μ_o and the optimum stabilizer constant μ_{so} do not alter noticeably, although the stability ratio s_o drops by about 40%, thus reducing the stability margin considerably.

When comparing a 2-pole non-condensing type of turbo-alternator with the 2-pole 30 MW condensing type, we find the inertia to be nearly 40% lower, whereas the time-constants are the same. It follows that the stability ratio should improve in the non-condensing alternator. The improvement applies to the high-gain limit only, since only H varies (see Fig. 5c).

The conclusions drawn for 30 MW alternators apply also if alternators of higher or lower ratings are compared.

(4.1.2) Water-Wheel Generators.

For water-wheel generators both H and τ'_{do} vary with speed. Considering, for example, a 30 MW 500 r.p.m. alternator and comparing it with a 100 r.p.m. alternator, we find that there is a drop of 25% in H and of 40% in τ'_{do} for the low-speed machine. The decrease in H hardly affects μ_o and μ_{so} , but increases the stability ratio by about 10%. However, the reduction in the time-constant reduces the stability ratio by about 20% and raises μ_{so} by 40%. It follows that the optimum operating conditions deteriorate for the lower-speed generators, and this applies to any output rating.

Comparing the water-wheel generators with a 3000 r.p.m. condensing-type turbo-alternator, we find that fast water-wheel generators are much more stable. Low-speed (100 r.p.m.) generators, however, provide a lower stability for ratings above approximately 30 MW.

(4.2) Comparison of Alternators of Different Output Rating

(4.2.1) Steam Turbo-Alternators.

In general, an increase in output capacity means a reduction in H and an increase in τ'_{do} for a steam turbo-alternator. For example, when doubling the capacity of a condensing-type 2-pole generator from 30 to 60 MW we find that there is an increase of 25% in τ'_{do} and a drop of 25% in H . These changes improve the stability markedly on a per-unit basis. The decrease in H affects only s_o , and increases it by roughly 10%. The increase in τ'_{do} increases s_o by at least another 70% and improves μ_o and μ_{so} as well. This means that the stability ranges are nearly doubled by the doubling of the output rating, and it follows, for example, that one single generator should provide much more stable operating conditions than a combination of two units of half the rating.

For other types of steam turbo-alternators similar improvements in the stability conditions arise with an increase in rating.

(4.2.2) Water-Wheel Generators.

For water-wheel generators the stability does not improve as markedly as for the steam turbo-alternator, since here the inertia constant H increases slightly with output rating. However, the relative increase in τ'_{do} is much greater, overriding the effects of H . Hence there is still a considerable overall increase in the dynamic stability ranges.

(5) CONCLUSION

The operation of an alternator under dynamic conditions is better for machines with lower per-unit inertia constant and larger field time-constant. This means that the relative dynamic stability of alternators can be determined by considering the magnitude of these two parameters if the alternator impedances do not vary.

Changes in alternator impedances also affect the dynamic stability limit and have to be allowed for in any general investigation.

When particular alternators are compared their relative dynamic stability can be deduced from Table 1 and Figs. 5A-5D. In general, it can be said that higher speed improves stability. In addition, the stability ranges improve considerably with an increase in output rating if they are compared on a per-unit basis.

(6) ACKNOWLEDGMENTS

The author would like to thank Prof. D. M. Myers for the use of the differential analyser of the C.S.I.R.O. at the University of Sydney. He also expresses his appreciation to Mr. J. R. Ward for assistance in the use of the electronic analogue computer of the Aeronautical Engineering Department of the University.

(7) REFERENCES

- (1) HEFFRON, W. G., and PHILLIPS, R. A.: "Effect of a Modern Amplidyne Voltage Regulator on Underexcited Operation of Large Turbine Generators," *Transactions of the American I.E.E.*, 1952, **71**, Part III, p. 692.
- (2) HEFFRON, W. G.: "A Simplified Approach to Steady-State Stability Limits," *ibid.*, 1954, **73**, Part IIIA, p. 39.
- (3) CONCORDIA, C.: "Effect of Buck-Boost Voltage Regulator on Steady-State Power Limit," *ibid.*, 1950, **69**, Part I, p. 380.
- (4) MESSERLE, H. K., and BRUCK, R. W.: "Steady-State Stability of Synchronous Generators as Affected by Regulators and Governors" (see page 24).
- (5) FIRST REPORT OF POWER SYSTEM STABILITY, A.I.E.E. Subcommittee on Interconnection and Stability Factors *Transactions of the American I.E.E.*, 1937, **56**, p. 261.
- (6) CRARY, S. B.: "Power System Stability," Vol. II (John Wiley and Sons, 1947).
- (7) CONCORDIA, C., KIRCHMAYER, L. K.: "Tie-Line Power and Frequency Control of Electric Power Systems," *Transactions of the American I.E.E.*, 1953, **72**, Part III, p. 562.
- (8) ZABORSZKY, J., and RITTENHOUSE, J. W.: "Electric Power Transmission" (The Ronald Press Co., 1954).

AN EXPERIMENTAL STUDY OF INTENSITY SPECTRA AFTER HALF-WAVE RECTIFICATION OF SIGNALS IN NOISE

By G. E. FELLOWS, Ph.D., M.S., B.A., and D. MIDDLETON, Ph.D., M.A., A.B.

(The paper was first received 17th June, and in revised form 16th September, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

Measurement has been made of the spectral intensity of the harmonic zones which exist at the output of a non-linear device fed by narrow-band noise and an unmodulated carrier. The spectral shape, the maximum spectral intensity, and the carrier component in each of the first six harmonic zones ($l = 0 \dots 5$) have been determined both theoretically and experimentally for a half-wave ν th-law rectifier with $\nu = 1/2, 1$ and 2 , for a wide range of input carrier/noise ratios.

The measuring equipment is discussed in the paper, the theoretical and experimental results are compared, and a number of computed results of practical interest are presented.

LIST OF PRINCIPAL SYMBOLS

- I = Current through detector load, amp.
- V = Voltage at detector input, volts.
- β = Detector dynamic transconductance, amp/volts ^{ν} .
- ν = Law of detector.
- f_0 = Frequency of input carrier, c/s.
- l = Order of harmonic zone.
- $\psi(\tau)$ = Correlation function of input noise voltage to detector.
- $R(\tau)$ = Correlation function of detector load current.
- p = Input carrier/noise power ratio.
- A_0 = Amplitude of input carrier, volts.
- Q = Q-factor of narrow-band filter.
- $W(f)$ = Intensity spectrum of detector output current (watts per cycle in a 1-ohm load resistor).
- $W_l(f)$ = Intensity spectrum of the l th zone.
- K_l = Intensity of the carrier component in zone l .
- p_l = Ratio of carrier component power to noise power in zone l .
- ω_F = Angular cut-off frequency of filter F_2 .

(1) INTRODUCTION

One of the more important phases of communication theory is the problem of the optimum extraction of signals from noise. Since most communication systems involve the use of some non-linear operation such as mixing or detection, and since the dynamic characteristic of the non-linear device provides a parameter in the optimization problem, the importance of a theory dealing with the passage of signals and noise through non-linear elements is apparent.

Techniques have been developed^{1,2,3,4} which are adequate for the solution of many non-linear problems of practical interest. The analytical results are all based on certain simplifying assumptions which make possible a tractable mathematical description of the noise in these non-linear elements. A characteristic feature of the theory for those mathematical models which adequately represent physical situations of importance is the complexity of some of the expressions for the output statistics.

Moreover, for some special cases of the general theory no way has been found to express the solution in computable form. The preceding remarks point to the need for an experimental approach to these problems on three closely related grounds. First, it is important to investigate experimentally a number of computable cases of the general theory in order to determine the adequacy of the assumptions used in the mathematical models. Secondly, by comparing these theoretical and experimental results we can establish the conditions under which the experimental results can be relied upon when problems lying beyond a possible analytical solution are investigated. Finally, actual results can be obtained experimentally for cases in which computation is difficult or impossible.

A number of experimental studies have been made^{5,6,7} which testify to the adequacy of the mathematical models in special instances. One particular case of the general theory has recently been investigated experimentally by Raymond and McLucas,⁸ indicating good qualitative agreement of theory with experiment. The purpose of the present study has been to investigate in more precise and extended fashion the problem examined by Raymond and McLucas, and to consider a number of more general cases in an effort to satisfy the three general objectives outlined above.

A half-wave ν th-law rectifier is defined by the characteristic

$$I = \begin{cases} \beta V^\nu; & V > 0 \\ 0; & V < 0 \end{cases} \quad \nu > 0$$

where V is the applied voltage, I is the output current, and β is a (dynamic) transconductance. The determination of the intensity spectrum at the output of a ν th-law rectifier fed by a carrier and narrow-band noise is a problem of practical importance, since the output of a well-tuned r.f. section or i.f. strip consists of narrow-band noise centred about the carrier frequency, and since the most commonly encountered detectors in a.m. communication systems are the half-wave linear ($\nu = 1$) and the half-wave quadratic ($\nu = 2$) elements.

The intensity spectrum at the output of a ν th-law rectifier consists of zones centred about harmonics of the central frequency of the input spectrum, as shown in Fig. 1. The l.f. and first-order zones are usually of greatest practical importance and have been carefully studied.^{9,10} In this paper, the properties of the first six spectral zones are discussed for $\nu = 1/2, 1$ and 2 , and for a wide range of input carrier/noise ratios. It is believed that the higher-order zones have not previously been investigated with this generality.

(2) THEORETICAL REMARKS

The correlation function of the output of a non-linear device is, in general, most easily obtained by the characteristic-function methods used by Rice and others.^{2,3} In this Section results are presented which have been obtained by applying this technique to the problem of half-wave rectification of narrow-band noise and an unmodulated carrier.

It is assumed that the narrow-band noise belongs to a

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Fellows is in the Research Department, Melpar Inc., Cambridge, Mass.

Dr. Middleton resides at Cambridge, Mass.

stationary, normal random process and is centred about a frequency f_0 . The carrier component $A_0 \cos 2\pi f_0 t$ and the noise are assumed to be statistically unrelated. The correlation function of the narrow-band noise can be expressed in the form

$$\psi(\tau) \equiv \psi_r(\tau) = \psi_{r0}(\tau) \cos \omega_0 \tau \quad (1)$$

where $r(\tau)$ is the normalized correlation function, and $r_0(\tau)$ is the normalized correlation function of the slowly-varying part of the narrow-band noise. The input carrier/noise power ratio p is then given by $p \equiv A_0^2/2\psi$.

Because of the zonal nature of the spectrum at the output of

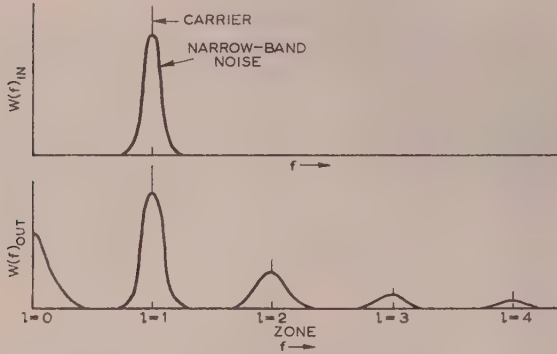


Fig. 1.—Typical intensity spectra at input and output of a non-linear element.

the non-linear device (cf. Fig. 1), the output correlation function can be expressed as

$$R(\tau) = \sum_{l=0}^{\infty} R_l(\tau) \quad (2)$$

where $R_l(\tau)$ is the correlation function of the harmonic zone located about the frequency $l f_0$, and the integer l indicates the order of the output zone.

For the case of half-wave ν th-law rectification of the inputs discussed above, solution by the characteristic-function method yields the result that*

$$R_l(\tau) = \cos l\omega_0\tau \sum_{m=0}^{\infty} [B_{|m-l|} + B_{m+l}(1-\delta\eta)] \quad (3)$$

where $\delta\eta$ is the familiar Kronecker delta, and

$$B_{|\alpha|} = \frac{\epsilon_m \beta^2 \Gamma^2(\nu+1) p^m l^\nu}{2^{\nu+2} (m!)^2} \sum_{q=0}^{\infty} \frac{r_0(\tau)^{2q+|\alpha|} {}_1F_2^2(\gamma; m+1; -p)}{q!(q+|\alpha|)! \Gamma^2(1-\gamma)} \quad (4)$$

Here α represents either $m+l$ or $m-l$, and $\gamma = (|\alpha| + 2q - \nu + m)/2$. The quantity ϵ_m is defined by $\epsilon_0 = 1$ and $\epsilon_m = 2$ for $m \geq 1$, and ${}_1F_1$ is the confluent hypergeometric function.¹¹

The narrow-band filter in the experimental equipment is shown in Fig. 2. For this circuit, a good approximation for sufficiently high Q -factor is given by

$$r_0(\tau) = \exp(-\omega_0|\tau|/2Q); Q^2 = R^2C/L \quad (5)$$

Substituting eqn. (5) in eqn. (4), we now employ the well-known Wiener-Khinchine theorem^{12,13} to obtain the output intensity spectrum $W(f)$:

$$W(f) = 4 \int_0^{\infty} R(\tau) \cos \omega \tau d\tau = \sum_{l=0}^{\infty} 4 \int_0^{\infty} R_l(\tau) \cos \omega \tau d\tau \equiv \sum_{l=0}^{\infty} W_l(f) \quad (6)$$

* Eqs. (3) and (4) are the results obtained by Middleton (Reference 3, Section 3), except that the term B_{m+l} in eqn. (3) was inadvertently omitted in the earlier paper.

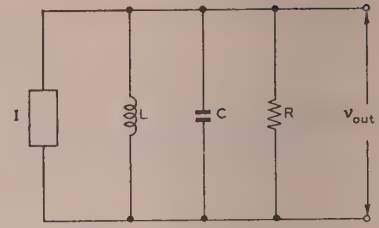


Fig. 2.—Narrow-band filter.

The d.c. level and the carrier components in the output spectrum appear as delta functions, and we may formally separate these components by writing

$$W_l(f) \equiv W_{l,cont}(f) + K_l \delta(f - l f_0) \quad (7)$$

so that the final expressions for the output zonal intensity spectra can now be written to a good approximation for sufficiently high Q -factors as

$$W_l(f)_{cont} = (1 + \delta\eta) \sum_{m=0}^{\infty} \frac{\epsilon_m p^m}{(m!)^2} \left[\sum_{q=0}^{\infty} \frac{D_{|m-l|}}{m+l} + (1 - \delta\eta) \sum_{q=0}^{\infty} \frac{D_{m+l}}{m+l} \right] \quad (8)$$

where

$$D_{|\alpha|} = \frac{Q \beta^2 \Gamma^2(\nu+1) \psi^\nu(|\alpha| + 2q) {}_1F_2^2(\gamma; m+1; -p)}{2^\nu \omega_0 q! (q+|\alpha|)! \Gamma^2(1-\gamma) [(|\alpha| + 2q)^2 + 4Q^2(\omega/\omega_0 - l)^2]} \quad (9)$$

$$\text{and } K_l = \frac{\beta^2 \Gamma^2(\nu+1) \psi^\nu p^l {}_1F_2^2[(l-\nu)/2; l+1; -p]}{2^{\nu+1} (l!)^2 \Gamma^2[1 - (l-\nu)/2]} \quad (10)$$

The total zonal intensity in the continuum is found by computing $[R_l(0) - R_l(\infty)]$ from eqn. (3). In Fig. 3 this quantity is shown

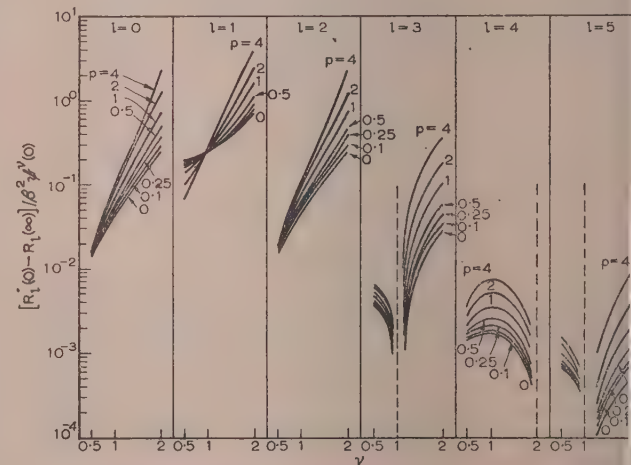


Fig. 3.—Calculated total zonal intensity in the continuum as a function of ν .

as a function of the detector law ν , with p as parameter. The carrier component K_l in each zone is shown in Fig. 4.*

We may define an output carrier/noise ratio p_l for each zone by the ratio $K_l/[R_l(0) - R_l(\infty)]$. The computed results show

* The spectral shape [from eqn. (8)], the total intensity of the continuum, and the carrier component K_l have been computed for $\nu = 1/2, 1$ and 2 . For each ν , the first six spectral zones have been determined over a wide range of input carrier/noise ratio ($p = 0; 0.1; 0.25; 0.5; 1; 2; 4; 10$). These computed results appear in Figs. 3 to 3.25 of Reference 14.

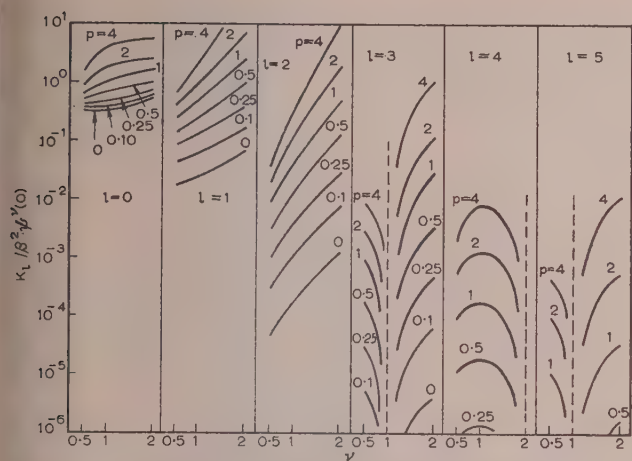


Fig. 4.—Calculated carrier components in the zones as a function of ν .

that p_l is very nearly independent of ν , and that the empirical relationship

$$\log_{10} p_l = \log_{10} p - 0.7(l - 1) \quad (11)$$

holds well for all zones computed, with the exception of $l = 0$.

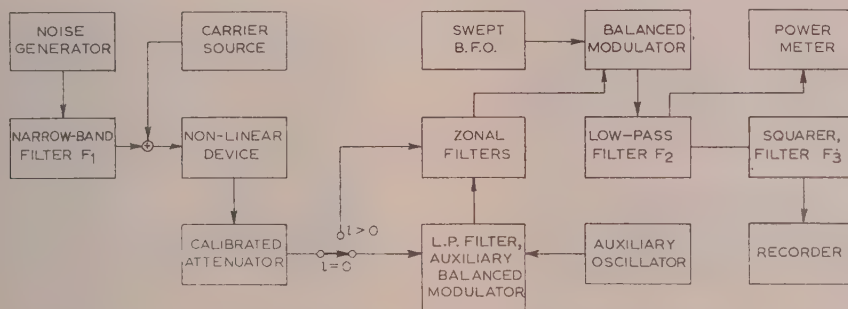


Fig. 5.—Block diagram of experimental equipment.

(3) EXPERIMENTAL PROCEDURES

A block diagram of the equipment used to measure $W_l(f)$ is shown in Fig. 5. An amplified shot-noise generator provides Gaussian noise with a spectrum flat to 2 Mc/s. The noise output of a narrow-band filter centred about 200 kc/s, and with a Q-factor of 25 is linearly added to a 200 kc/s carrier and sent to the non-linear device. An input carrier/noise ratio of unity is then obtained by adjustment of the carrier amplitude to give the same reading on a thermal meter as is given by the narrow-band noise. Other carrier/noise ratios follow by an appropriate setting of a calibrated attenuator in the carrier source. For the study of all zones except the low-frequency zone, the output of the non-linear device is sent through a high-pass filter which eliminates all zones lower in frequency than the one being measured. Sufficient amplification is provided to maintain the r.m.s. voltage into the balanced modulator nearly constant for all zones. The latter two operations are made necessary by the rapid decrease in zonal power with the order, l , of the zone. The swept beat-frequency oscillator (b.f.o.) provides an output which scans slowly from a frequency $(l + 0.2)f_0$ to a frequency lf_0 . The output of this unit is applied to the balanced modulator, which is adjusted to yield an output spectrum consisting of the input spectrum shifted downward by an amount equal to the swept b.f.o. frequency. Since the balanced modulator output is passed

through a very sharp cut-off low-pass filter (90 dB/octave) with a cut-off frequency of only $0.0025f_0$, the intensity at the output of this filter as a function of scan time is just the intensity spectrum of the zone under investigation.

The low-frequency zone is handled in the same way, except that it is first modulated *upward* in frequency in the auxiliary balanced-modulator, so that it has the spectral appearance of one of the higher-order zones.

The intensity at the output of the low-pass filter is measured by applying the output voltage of this filter to an analogue squarer.* The squarer output is then smoothed and sent to a recorder such as the Sanborn Model 127.

The only block requiring special attention here is the non-linear device. This unit consists of a diode function-generator,† containing two banks of four parallel linear-detectors each. The gain C_k and the bias b_k of each of the eight detectors are independently adjustable. By adding the voltage outputs of the two banks, half-wave characteristics with positive curvature can be synthesized as shown in Fig. 6(a). If the voltage output of the bank containing detectors 5–8 is subtracted from the output of the bank containing detectors 1–4, half-wave characteristics with negative curvature can be formed as shown in Fig. 6(b). The photographs (from an oscillographic presentation of these characteristics) are examples of the accuracy obtainable with this circuit in each mode of operation.

The design parameters of the total system were selected to

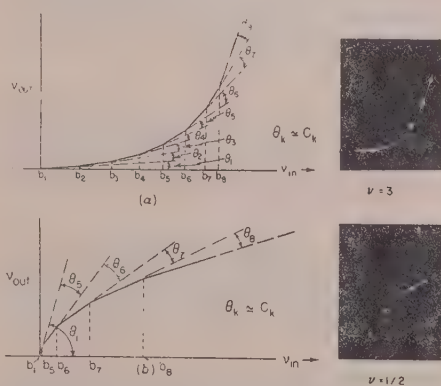


Fig. 6.—Typical synthesis of non-linear characteristics.

(a) $\nu = 3$. (b) $\nu = \frac{1}{2}$.

maintain the sum of the statistical and system errors at less than 2%, relative to the maximum intensity in a given zone. The output of any averaging device which acts upon a random time

* The squarer used here is a modification of a multiplier originally developed by Dr. Peter Elias (Cruft Laboratory, Harvard), and used by Johnson (see Reference 7).

† Diode function-generators have found wide application as non-linear elements in analogue computers. A general discussion of these devices appears in Reference 15, along with the specific details of the unit mentioned here.

series for a finite time will fluctuate about the average value. The ratio of the r.m.s. fluctuation to the average value of the wave is termed the statistical error. In measuring the average intensity in a random wave which has a spectrum extending from direct current to an angular frequency ω_F , it has been shown¹⁶ that the statistical error is a function of ω_F , the averaging time T , and the time-constant of the averaging device. However, if the averaging time exceeds the time-constant of the averaging device (as is usually the case), the statistical error approaches a minimum independent of T . As applied here, the averaging element is a low-pass RC filter and ω_F is the angular cut-off frequency of filter F_2 . For this case the minimum statistical error is approximately equal to $(RC\omega_F)^{-1/2}$. An increase in the time-constant RC to reduce the statistical error results in an increase in the scanning error, which arises from the fact that a non-zero rate of sweep introduces filter transient errors in the subsequent measurements. Similarly, an increase in ω_F results in the increase of another system error, since the intensity at the output of filter F_2 is proportional to

$$\int_{f-\omega_F/2\pi}^{f+\omega_F/2\pi} W_l(f') df'$$

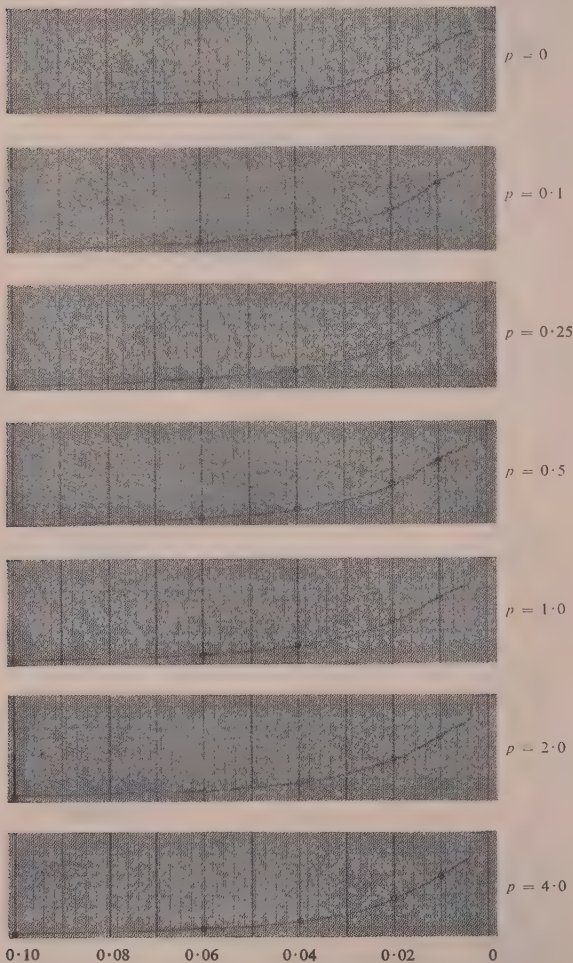


Fig. 7.—Measured intensity spectrum of zone $l = 1$, after half-wave square-law ($\nu = 2$) detection of narrow-band noise plus an unmodulated carrier.

Theoretical points normalized to agree with experimental value at $(\omega/\omega_0 - 1) = 0.01$.

and this quantity approximates to $W_l(f)$ only if ω_F is very small compared to the spectral width of the zone under investigation. All these errors can be made sufficiently small either by increasing the centre frequency f_0 of the input noise or by increasing the scan time. The latter alternative, however, puts a strain on system stability. By setting the scan time at a reasonable maximum of 10 min, we obtain a total relative error of less than 2% by increasing the frequency f_0 to 200 kc/s and choosing a value for ω_F of 3600 rad/sec. Representative sets of experimentally obtained spectra are shown in Figs. 7 ($\nu = 2$; $l = 1$), 8 ($\nu = 1$; $l = 4$) and 9 ($\nu = \frac{1}{2}$; $l = 0$).

The zonal intensity spectra can be completely specified by measuring the spectral shape, the intensity of the carrier component, and either the maximum intensity or the total intensity in the continuum. The latter has the advantage of a simpler interpretation of results, but the measurement of the maximum zonal intensity can be accomplished with very little additional equipment over that required for measurement of the spectral distribution. The zonal intensity measurements were carried out, in fact, by simply sending the output of filter F_2 to a carefully

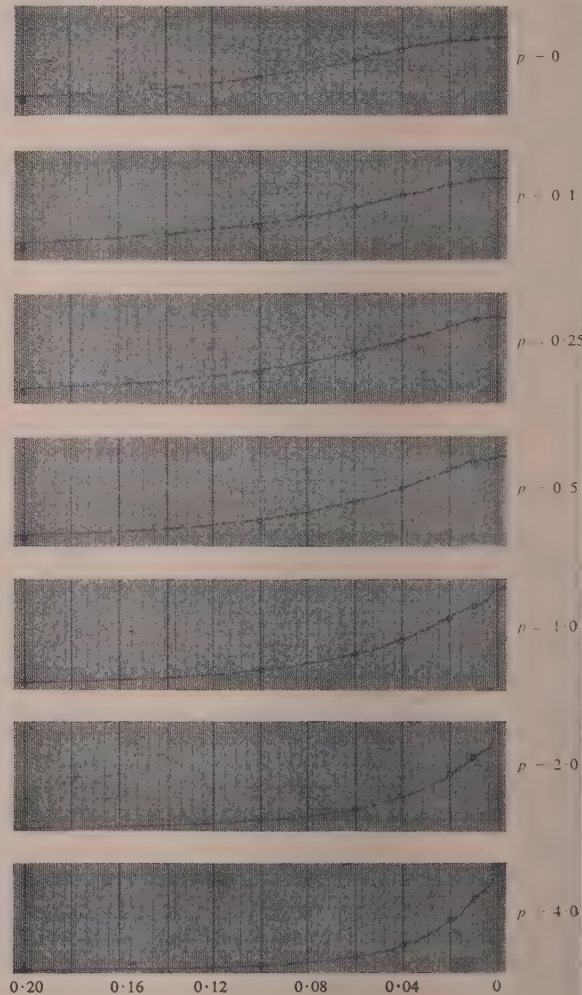


Fig. 8.—Measured intensity spectrum of zone $l = 4$, after half-wave linear ($\nu = 1$) detection of narrow-band noise plus an unmodulated carrier.

Theoretical points normalized to agree with experimental value at $(\omega/\omega_0 - 4) = 0.01$.

calibrated thermal meter and by adjusting the b.f.o. to two successive fixed frequencies for each zone, as discussed below.

If we designate the b.f.o. output frequency by f_1 and let $f' = f - f_1$, the mean-square voltage \bar{v}_0^2 at the output of filter F_2 may be written as a function of f_1 in the form

$$\bar{v}_0^2(f_1) = G^2 \int_{-\infty}^{\infty} |H(f')|^2 W_l(f_1 + f') df' \quad (12)$$

where G is the product of the gains of the zonal filter, the conversion gain of the balanced modulator, and the pass-band gain of filter F_2 . $H(f')$ is the relative transfer function of filter F_2 , with a value of unity in the pass band, and $H(f')$ equals $H(-f')$. Since $W_l(f_1 + f')$ is nearly constant for values of f' for which $H(f')$ is different from zero, and since we may express $W_l(f)$ as $W_l(f)_{cont} + K_l \delta(f - lf_0)$, we may write

$$\bar{v}_0^2(f_1) = G^2 W_l(f_1)_{cont} \int_{-\infty}^{\infty} |H(f')|^2 df' + G^2 K_l \int_{-\infty}^{\infty} \delta(f' + f_1 - lf_0) |H(f')|^2 df' \quad (13)$$

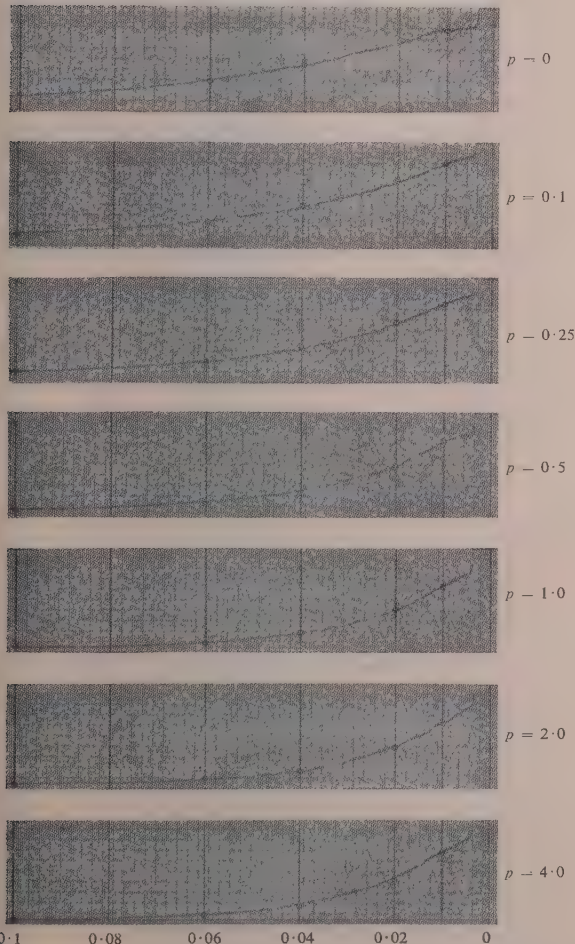


Fig. 9.—Measured intensity spectrum of zone $l = 0$, after half-wave detection ($\nu = 1/2$) of narrow-band noise plus an unmodulated carrier.

Theoretical points normalized to agree with experimental value at $(\omega/\omega_0) = 0.01$.

The integral of $|H(f')|^2$ has been carefully determined by graphical integration to be 1240 for the system in use here. Because of the extremely sharp cut-off characteristic of filter F_2 beyond the frequency $0.0025f_0$, very little error is introduced by writing

$$\bar{v}_0^2(f_0) \doteq 1240 G^2 W_l(lf_0)_{cont} + G^2 K_l \quad (14)$$

and

$$\bar{v}_0^2[(l + 0.01)f_0] = 1240 G^2 W_l[(l + 0.01)f_0]_{cont} = 1240 \eta G^2 W_l(lf_0)_{cont} \quad (15)$$

Here η is the theoretically determined ratio

$$W_l[(l + 0.01)f_0]_{cont} / W_l(lf_0)_{cont}$$

In the absence of theoretical data, η can easily be obtained by extrapolating the experimental spectral curves (Figs. 7, 8 or 9, for example).

Our object here is to obtain absolute measurements in the sense that no normalization is needed for the comparison of the experimental and theoretical results. Since the theoretical results are expressed in terms of the quantity $W_l(f)\omega_0/Q\beta^2\psi$, it appears necessary that the constants Q , β , ω_0 , ψ and G must be accurately determined. We are able to eliminate this step, however, by the following technique: Consider the output of filter F_2 when the non-linear device is converted to a linear amplifier with gain β , and is fed by noise alone. In the usual fashion, by linear circuit theory, it can be shown that

$$\bar{v}_0^2(f_0)_{lin} = \frac{4 \times 1240 G^2 \beta^2 \psi Q}{\omega_0} \quad (16)$$

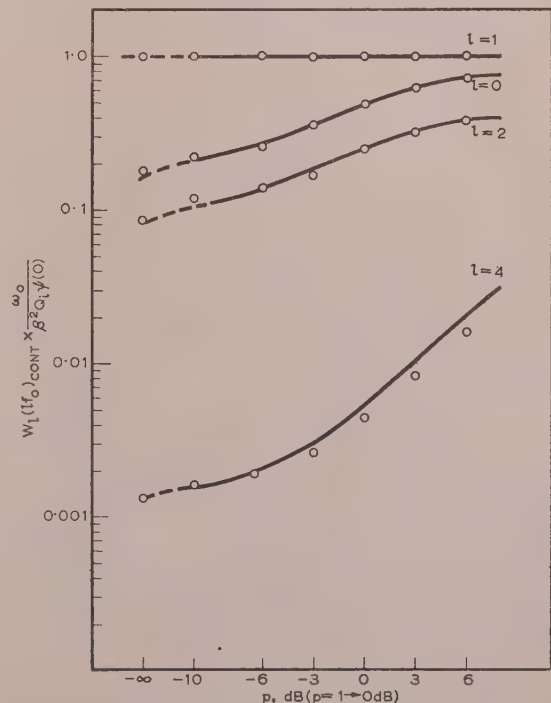


Fig. 10.—Measured peak zonal spectral intensity in the continuum as a function of the carrier/noise ratio p at the input to a half-wave linear ($\nu = 1$) detector.

where the subscript *lin* labels this particular measurement. Next, we return to eqns. (14) and (15), and use eqn. (16) to write

$$W_l(lf_0)_{cont} \frac{\omega_0}{Q\beta^2\psi} = 4\eta \frac{\bar{v}_0^2[(l+0.01)f_0]}{\bar{v}_0^2(f_0)_{lin}} \quad (17)$$

$$\text{and } K_l \frac{\omega_0}{Q\beta^2\psi} = \frac{4 \times 1240}{\bar{v}_0^2(f_0)_{lin}} \{ \bar{v}_0^2(lf_0) - \eta \bar{v}_0^2[(l+0.01)f_0] \} \quad (18)$$

It is understood that all the constants listed above are left unchanged during all measurements, including that of $\bar{v}_0^2(f_0)_{lin}$. Examples of the results obtained by this technique are shown in Figs. 10 and 11 for the case $\nu = 1$.

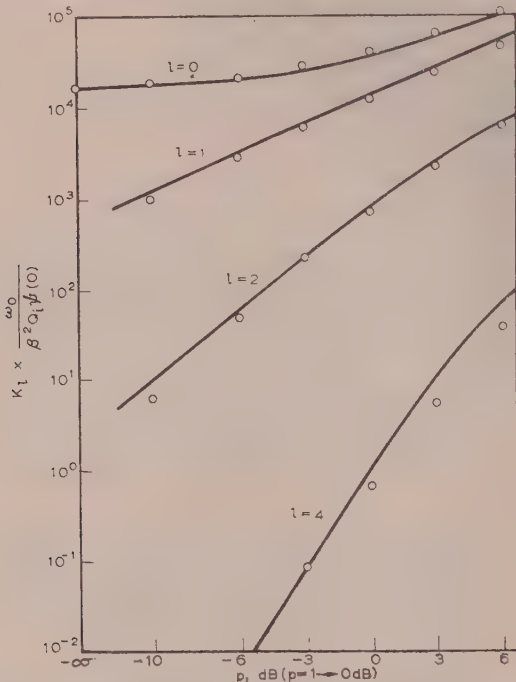


Fig. 11.—Measured carrier components in the output of a half-wave linear ($\nu = 1$) detector as a function of the input carrier/noise ratio p .

(4) RESULTS* AND CONCLUSIONS

In Figs. 7, 8 and 9, theoretical points have been plotted on the recordings of spectral shape, after normalization at the point $(\omega/\omega_0 - l) = 0.01$. The relative error never exceeds 2% for these sample recordings, but tends to increase slightly for the higher-order zones. Theoretical curves are also plotted on Figs. 10 and 11 for comparison with the zonal intensity measurements. Although the results cover a range of about 60 dB, at only one or two points does the discrepancy between theory and experiment exceed 2 dB. Measurements for the case of $\nu = 1$ proved to be the most accurate because of the relative ease in setting up the non-linear device for this case. The results for $\nu = \frac{1}{2}$ and 2 are only slightly inferior, however, with an error seldom greater than 3 dB, and then only for the higher-order zones. In all cases, the error in the measurement of a given zone is in the same direction for all values of p , resulting in the preservation of the "shape" of the curves.

In an attempt to account for the small errors mentioned above, data on spectral shape and zonal intensity were obtained as a function of the detector bias. It was found that a very small

error in the bias setting could easily cause the observed errors in both spectral shape and intensity.

In general, the agreement of theory with experiment appears to warrant a high degree of confidence in those portions of the general theory which have been studied here experimentally. In addition, the experimental conditions have been established which will permit investigation of a number of other special cases of interest with known experimental accuracy.

(5) ACKNOWLEDGMENTS

The research reported here was made possible through support extended to the Cruft Laboratory, Harvard University, through Office of Naval Research Contract N5ori-76 (T.O.1), jointly with the Signal Corps, Air Material Command, and ONR.

(6) REFERENCES

- (1) BENNETT, W. R.: "The Response of a Linear Rectifier to Signal and Noise," *Journal of the Acoustical Society of America*, 1944, **15**, p. 164.
- (2) RICE, S. O.: "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, 1944, **23**, p. 282, and 1945 **24**, p. 46.
- (3) MIDDLETON, D.: "Some General Results in the Theory of Noise through Non-Linear Devices," *Quarterly of Applied Mathematics*, 1948, **5**, p. 445.
- (4) NORTH, D. O.: "The Modification of Noise by Certain Non-Linear Devices," Paper read before the Institute of Radio Engineers, January 28, 1944.
- (5) KNUDTZON, N.: "Experimental Study of Statistical Characteristics of Filtered Random Noise," Research Laboratory of Electronics (Massachusetts Institute of Technology), Technical Report 115, 1949.
- (6) LEE, Y. W., and STUTT, C. A.: "Statistical Prediction of Noise," Research Laboratory of Electronics (Massachusetts Institute of Technology), Technical Report 129, 1949.
- (7) JOHNSON, R. A.: "The Rectification of Signals and Random Noise by a Linear Detector," Cruft Laboratory (Harvard University), Technical Reports 144 and 145, 1952.
- (8) RAYMOND, R. C., and MCLUCAS, J. L.: *Journal of Applied Physics*, 1951, **22**, p. 1211.
- (9) MIDDLETON, D.: "Rectification of a Sinusoidally Modulated Carrier in the Presence of Noise," *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 1467.
- (10) DAVENPORT, W. B.: "Signal-to-Noise Ratios in Bandpass Limiters," Research Laboratory of Electronics (Massachusetts Institute of Technology), Technical Report 23, 1952. Also *Journal of Applied Physics*, 1953, **24**, p. 720.
- (11) MIDDLETON, D., and JOHNSON, V.: "A Tabulation of Selected Confluent Hypergeometric Functions," Cruft Laboratory (Harvard University), Technical Report 140, 1952.
- (12) WIENER, N.: "Generalized Harmonic Analysis," *Acta Mathematica*, 1930, **55**, p. 117.
- (13) KHINTCHINE, A.: *Mathematische Annalen*, 1934, **109**, p. 60.
- (14) FELLOWS, G. E., and MIDDLETON, D.: "Intensity Spectra after Half-Wave Detection of Signals in Noise," Cruft Laboratory (Harvard University), Technical Report 21, 1955.
- (15) FELLOWS, G. E.: "Intensity Spectra after Half-Wave Detection of Signals in Noise," Cruft Laboratory (Harvard University), Technical Report 218, 1955.
- (16) DAVENPORT, W. B., JOHNSON, R. A., and MIDDLETON, D.: "Statistical Errors in Measurements on Random Time Functions," *Journal of Applied Physics*, 1952, **23**, p. 37.

* The complete set of spectra shape and intensity measurements for $\nu = \frac{1}{2}$, 1 and 2 appears in Reference 15, Section 5, Figs. 5.1 to 5.21.

ON THE INDUCTANCE OF IRON-CORED COILS

By P. HAMMOND, M.A., Associate Member.

(The paper was first received 9th June, 1955, and in revised form 26th September, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

The magnetic field in the neighbourhood of coils threaded through holes in iron cylinders is calculated. The results are relevant to the estimation of the self-inductance of rotating machines and transformers and form the basis for an examination of the soundness of approximate formulae used in the design of such apparatus.

(1) INTRODUCTION

In many electrical devices it is important to be able to estimate the inductance of current-carrying coils, which are threaded through holes in iron. The problem arises with different types of rotating machines and also with static devices such as transformers and reactors. As a result there is a great deal of design experience, and approximate formulae are used successfully every day. In spite of this success, however, it may be of considerable interest to examine in the light of magnetic theory some of the assumptions that underlie the formulae that are used. This is the object of the paper. Since, in general, exact solutions of particular examples of iron-cored coils are impossible to obtain, at least without immense mathematical labour, we shall examine certain idealized cases. Three simplifying restrictions will be employed. First, we shall restrict ourselves to two-dimensional solutions. If, for instance, the problem is thought of as that of a rotating machine, this means that the effects of the end-connections will be ignored. Secondly we shall deal only with coils carrying steady currents. The solutions obtained will, however, apply substantially to alternating currents at power frequencies, i.e. to the problems referred to as "quasi-steady-state" in the textbooks. Eddy-current and hysteresis effects are outside the scope of the paper. Thirdly we shall deal only with iron cores that are circular cylinders.

(2) THE MAGNETIC FIELD OF CERTAIN SIMPLE COILS THREADED THROUGH HOLES IN AN IRON CYLINDER

(2.1) Single-Turn Coil inside Iron Cylinder

The problem is illustrated by Fig. 1, from which it will be seen that the iron cylinder has radius R and that the coil is threaded through small holes at a distance c from the axis of the cylinder.

If it were desired to find the self-inductance of this coil, the average magnetic force around a path such as is shown by the dotted line in the Figure might be calculated. It is then tempting to assume that this force is more or less constant between the conductors, and this presents a quick way of estimating the flux linked and hence the self-inductance. Such a procedure would, of course, never be employed for the calculation of the inductance of a single-turn coil in the absence of iron, because it is well known that the magnetic force varies considerably with the distance from the conductors. The question therefore arises as to what is the action of the iron on the magnetic field between the conductors. If this action causes the magnetic field to be more nearly uniform between the conductors, the approximate method

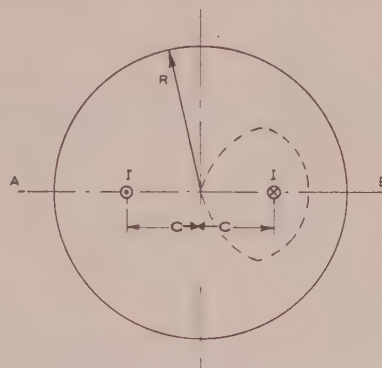


Fig. 1.—Single-turn coil threaded through small holes inside a solid iron cylinder.

outlined above can be justified. If not, the approximate method will give incorrect results.

The iron alters the magnetic field, because it will itself be magnetized by the coil. More explicitly, polarity will be induced on the surface of the iron, as shown in Fig. 2.

The amount and distribution of this polarity have been

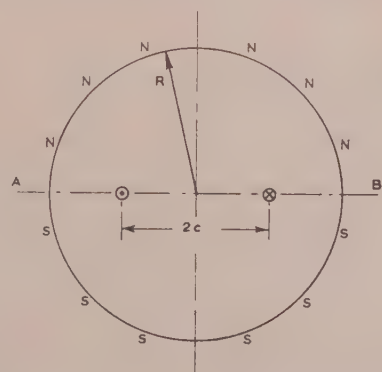


Fig. 2.—Surface polarity induced on an iron cylinder.

calculated in Section 8, and they are found to be almost independent of the permeability of the iron. This remarkable fact enables results to be estimated with considerable accuracy for iron of very varying quality and in different states of saturation. The reason for this is that the permeability μ enters into the expressions for surface polarity only in the ratio $(\mu - 1)/(\mu + 1)$ and this is always close to unity and therefore independent of μ .

Consider then the magnetic field between the conductors on the diameter passing through them. It is given at a distance r from the centre of the cylinder (see Section 8.5.1) by the expression

$$\frac{H_0}{4I} = \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots \right) - \left(\frac{c}{R^2} + \frac{c^3 r^2}{R^6} + \frac{c^5 r^4}{R^{10}} + \dots \right) \quad (1)$$

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Hammond is in the Department of Engineering, University of Cambridge.

The first infinite series gives the field due to the currents and the second that due to the surface polarity on the cylinder. Consider now some typical values of the ratio c/R , i.e. of the position of the conductors in the cylinder.

If $c/R = 1/2$ then

$$\frac{H_0}{4I} = \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) - \left(\frac{1}{4c} + \frac{r^2}{64c^3} + \frac{r^4}{1024c^5} + \dots\right)$$

The second series converges very rapidly and it is sufficient to take account of the first term, which is the contribution of the fundamental ($\sin \theta$) term of the surface polarity [see eqn. (21)].

Thus
$$\frac{H_0}{4I} \approx \left(\frac{3}{4c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

Fig. 3 shows a graph of this field, and it will be seen that the action of the iron by no means results in a constant magnetic

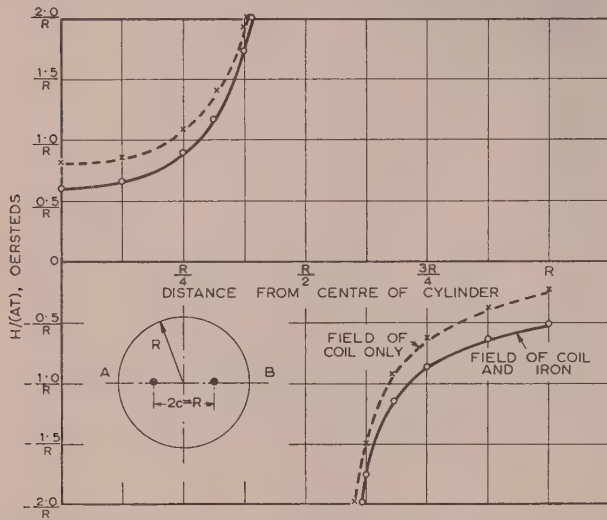


Fig. 3.—Field-form curve of a single-turn coil inside an iron cylinder.

force, rather does this action increase the variation of this force. Thus a method of estimating the inductance which assumes constant magnetic force is wrong in principle. The inductance of this coil is practically the same as if it were immersed in an infinite mass of iron. The correct procedure for estimating the inductance is first the calculation of the magnetic force in the absence of iron, then the calculation of flux density by means of the relevant B/H curve. Thence the flux and the inductance can be found.

If $c/R \rightarrow 1$, i.e. if the coil is threaded through holes in the iron very close to the surface of the cylinder, the magnetic field between the conductors is given as before by

$$\frac{H_0}{4I} = \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) - \left(\frac{c}{R^2} + \frac{c^3r^2}{R^6} + \frac{c^5r^4}{R^{10}} + \dots\right) \quad (1)$$

Therefore

$$\frac{H_0}{4I} \rightarrow \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) - \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

Therefore
$$\frac{H_0}{4I} \rightarrow 0$$

This is a very surprising result. The surface polarity in the previous case of $c/R = 1/2$ hardly affected the field distribution, whereas in this case it seems to have entirely cancelled out the magnetic field of the coil. It is desirable to investigate this a

little closer to see whether the field is actually zero or whether it is only approximately so. It must be remembered that the expression for H_0 is derived on the assumption that $(\mu - 1)/(\mu + 1) = 1$ for the iron. Without this assumption the expression for the magnetic field becomes

$$\frac{H_0}{4I} = \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) - \frac{\mu - 1}{\mu + 1} \left(\frac{c}{R^2} + \frac{c^3r^2}{R^6} + \frac{c^5r^4}{R^{10}} + \dots\right) \quad (2)$$

If $\frac{c}{R} \rightarrow 1$,
$$\frac{H_0}{4I} \rightarrow \left(1 - \frac{\mu - 1}{\mu + 1}\right) \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

Therefore
$$\frac{H_0}{4I} \rightarrow \frac{2}{\mu + 1} \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) \approx \frac{2}{\mu} \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

Thus the magnetic force has been reduced to $2/\mu$ times the value it would have in the absence of the iron.

Consider now the flux density B_0

$$\frac{B_0}{4I} = \frac{\mu H_0}{4I} \approx 2 \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) \quad (3a)$$

Thus the flux density has been doubled by the effect of the iron, and we have arrived at the interesting and useful conclusion that the presence of the iron cylinder will effectively double the self-inductance of such a coil.

In the paper we are concerned chiefly with coils threaded through holes in iron, but for completeness it is desirable to include the case of a coil placed on the surface of an iron cylinder, as shown in Fig. 4. This case is of great interest because it is in essence that of an armature coil placed in open slots.

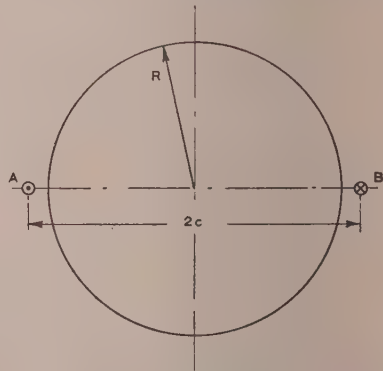


Fig. 4.—Single-turn coil outside an iron cylinder.

The magnetic force is calculated in Section 8.5.2 and is given by the expression

$$\frac{H_0}{4I} = \left(1 - \frac{\mu - 1}{\mu + 1}\right) \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

Therefore
$$\frac{B_0}{4I} = 2 \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) \quad (3b)$$

Thus for this case, as for the last, the iron cylinder doubles the self-inductance.

From a consideration of these three cases it will be seen that if the coil is placed within the iron and at some considerable distance from its surface, the self-inductance will be increased in the ratio of the order of the permeability, μ , neglecting saturation. Close to the surface of the cylinder, however, the self-inductance

of the coil, whether in open or closed slots, is merely doubled—a result that is substantially independent of the quality of the iron.

(2.2) Coil of Two Turns inside Iron Cylinder

In Section 2.1 we obtained some general insight into the effect of the induced polarity of the iron cylinder on the self-inductance of coils threaded through holes in the iron. It is now desirable to examine in greater detail the self-inductance of coils placed well within the iron. Consider the case illustrated in Fig. 5,

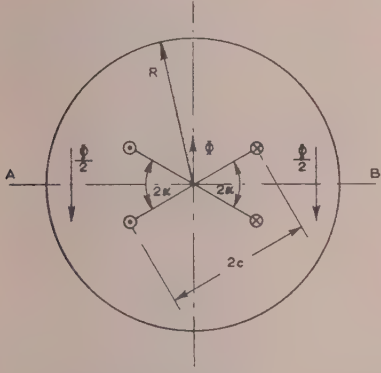


Fig. 5.—Two-turn coil inside an iron cylinder.

where $c = R/2$, so that the return path of the flux has a reluctance of the same order as the flux path between the conductors. The case under consideration is then not unlike that of a shell-transformer.

It was pointed out in Section 2.1 that the chief obstacle in the way of a reliable approximate solution was the large variation of magnetic field strength across the diameter of the cylinder. This variation made it unlikely that a solution, using the circuital law $\oint H dl = \frac{4\pi}{10}$ ampere-turns and assuming a constant value of H , would be valid. However, such solutions are used with considerable success. This suggests that the magnetic field strength must vary far less than is shown in Fig. 3. We need to examine the reason for this constancy of magnetic field strength, and as a step towards such an examination it is proposed to examine the problem of a two-turn coil as shown in Fig. 5.

Let the angle subtended at the centre of the cylinder by two conductors carrying current in the same direction be 2α . Then the magnetic force at the diameter AB due to the conductors and to the polarity of the cylinder, at distances from the centre of r less than c , is given in Section 8.5.3 as

$$\frac{H_0}{8I} = \left(\frac{\cos \alpha}{c} + \frac{r^2}{c^3} \cos 3\alpha + \frac{r^4}{c^5} \cos 5\alpha + \dots \right) - \left(\frac{c}{R^2} \cos \alpha + \frac{c^3 r^2}{R^6} \cos 3\alpha + \frac{c^5 r^4}{R^{10}} \cos 5\alpha + \dots \right) \quad (4)$$

Inspection of this expression reveals that it is very similar to eqn. (1), except that every term has been multiplied by a corresponding cosine. Since we are taking $c = R/2$, we can write the expression as

$$\frac{H_0}{8I} = \left(\frac{\cos \alpha}{c} + \frac{r^2 \cos 3\alpha}{c^3} + \frac{r^4 \cos 5\alpha}{c^5} + \dots \right) - \left(\frac{\cos \alpha}{4c} + \frac{r^2 \cos 3\alpha}{64c^3} + \frac{r^4 \cos 5\alpha}{1024c^5} + \dots \right)$$

The second part of this expression, which gives the field of the surface polarity, will be seen to converge even more rapidly than before. Once again it will be true to say that the field due to

this polarity will be largely independent of r and will therefore be uniform across the diameter of the cylinder. Since it will reduce the field between the conductors, the surface polarity will slightly increase the variation of magnetic field strength in this region. But the effect is bound to be small for conductors placed well within the iron. Thus, while it should be noted that the surface polarity cannot under any conditions cause uniformity of field strength between the conductors, its effect can be neglected so long as c/R is appreciably less than unity.

We shall therefore combine the two series of terms in eqn. (4) and write

$$\frac{H_0}{8I} = \frac{3 \cos \alpha}{4c} + \frac{r^2 \cos 3\alpha}{c^3} + \frac{r^4 \cos 5\alpha}{c^5} + \dots$$

It is clear that the terms containing r will be greatly reduced by their cosine multipliers. Thus we have here a mechanism that will render the field far more uniform. Fig. 6 shows the varia-

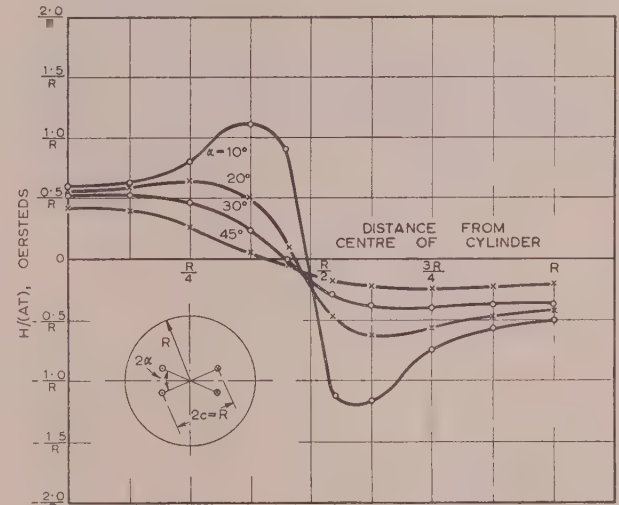


Fig. 6.—Field-form curves of two-turn coils inside iron cylinders.

tion of magnetic force across the diameter AB for various values of the angle α . The effect is very marked indeed and is similar to the effect of the winding factors in rotating machines. The process of distributing the winding results in a field, in which the fundamental component predominates and the harmonics are reduced in size. A completely uniform field would result from a sinusoidal distribution of current and in this case we should have only the first term of the Fourier series for the field. Such a winding would be awkward to construct. It is therefore all the more noteworthy that a winding of only two turns can result in a field which is substantially uniform over three-quarters of the distance between the "go" and "return" conductors.

(2.3) Consideration of the Magnetic Field for $r > c$

In Sections 2.1 and 2.2 attention was confined to the magnetic field for values of $r < c$, in order to simplify the presentation and because a knowledge of the flux crossing the diameter within that region is, of course, sufficient for a calculation of inductance. But Figs. 3 and 6 show the complete field-form curves for various coils, and we now briefly give the appropriate expressions for the field in the region $c < r < R$. For a single-turn coil (Figs. 1 and 2) it is shown in Section 7.5.1 that

$$\frac{H_0}{4I} = \left(\frac{c}{r^2} + \frac{c^3}{r^4} + \frac{c^5}{r^6} + \dots \right) + \left(\frac{c}{R^2} + \frac{c^3 r^2}{R^6} + \frac{c^5 r^4}{R^{10}} + \dots \right) \quad (5)$$

Hence if $c = R/2$

$$-\frac{H_0}{4I} = \left(\frac{c}{r^2} + \frac{c^3}{r^4} + \frac{c^5}{r^6} + \dots \right) + \left(\frac{1}{4c} + \frac{r^2}{64c^3} + \frac{r^4}{1024c^5} + \dots \right)$$

For a two-turn coil the expressions are

$$-\frac{H_0}{8I} = \left(\frac{c}{r^2} \cos \alpha + \frac{c^3}{r^4} \cos 3\alpha + \frac{c^5}{r^6} \cos 5\alpha + \dots \right) + \left(\frac{c}{R^2} \cos \alpha + \frac{c^3 r^2}{R^6} \cos 3\alpha + \dots \right) \quad (6)$$

and if $c = R/2$

$$-\frac{H_0}{8I} = \left(\frac{c}{r^2} \cos \alpha + \frac{c^3}{r^4} \cos 3\alpha + \dots \right) + \left(\frac{\cos \alpha}{4c} + \frac{r^2 \cos 3\alpha}{64c^3} + \dots \right)$$

It should perhaps be pointed out explicitly that in the region $c < r < R$ the field of the surface polarity adds to the field of the coils and thereby renders H_0 more nearly uniform. Table 1 gives values for the field, arising from the surface polarity only, for various two-turn coils spaced at $c = R/2$. The Table emphasizes the remarkable constancy of this component of the field.

Table 1
FIELD FOR TWO-TURN COILS

$\frac{r}{R}$	H_0 per ampere turn			
	$\alpha = 10^\circ$	$\alpha = 20^\circ$	$\alpha = 30^\circ$	$\alpha = 45^\circ$
0.55	$\frac{0.21}{R}$	$\frac{0.19}{R}$	$\frac{0.18}{R}$	$\frac{0.13}{R}$
0.625	$\frac{0.22}{R}$	$\frac{0.2}{R}$	$\frac{0.17}{R}$	$\frac{0.13}{R}$
0.75	$\frac{0.22}{R}$	$\frac{0.2}{R}$	$\frac{0.17}{R}$	$\frac{0.12}{R}$
0.875	$\frac{0.23}{R}$	$\frac{0.2}{R}$	$\frac{0.17}{R}$	$\frac{0.11}{R}$
1.0	$\frac{0.25}{R}$	$\frac{0.21}{R}$	$\frac{0.16}{R}$	$\frac{0.1}{R}$

(2.4) Field of Multi-turn Coils

In principle we have now solved the problem of the determination of the magnetic field, and hence of the inductance, for coils containing any number of turns threaded through holes anywhere within an iron cylinder. The resultant field can always be obtained by superposition, since we have shown that the magnetic force does not appreciably depend on the permeability of the iron. Fig. 7 shows the field-form curve of a five-turn coil, for which $c = R/2$. It will be seen that the magnetic force is far more uniform than that of Fig. 3. Close to the conductors which lie on the diameter AB the field does, however, increase very rapidly. This example is merely given as a typical case of a multi-turn coil.

(3) MAGNETIC FIELD OUTSIDE THE IRON CYLINDER

It was assumed in Fig. 5 that the flux Φ linking the coil would divide into two equal parts $\Phi/2$ and that this flux would remain within the iron. This, however, is clearly impossible.¹ In the calculations, we have made much use of the effect of the surface

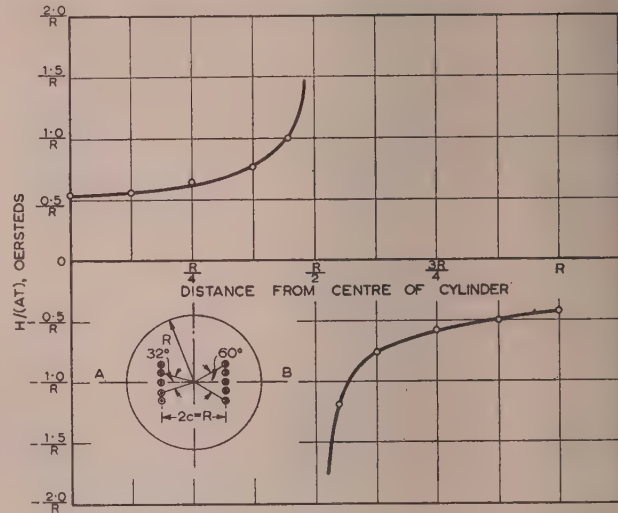


Fig. 7.—Field-form curve of a five-turn coil inside an iron cylinder

polarity on the cylinder. But surface polarity implies that flux is emerging from the iron into the air. This might be termed the leakage flux and it is desirable to know its order of magnitude.

Clearly the amount of this leakage flux will be affected by the value of c/R . For the cases where $c/R \rightarrow 1$ (see Section 2.1) practically all the return flux has to leave the iron. This is one explanation of the fact that the inductance of such coils is only doubled by the presence of the iron, instead of being increased in the ratio of the permeability. What has happened is that the entire return path of the flux is in air, and so the reluctance of the flux path has only been halved.

Consider now the case where $c/R = 1/2$, i.e. the case of conductors deeply buried in the iron. It is not immediately obvious what the leakage flux will be, and so it must be obtained from a consideration of the surface polarity. For the sake of simplicity we shall discuss the case of a single-turn coil, as in Figs. 1 and 2.

The leakage flux is given in Section 8.5.4 by the expression

$$\Phi_L = \frac{2\mu}{\mu + 1} \int_0^\pi H_r R d\theta \quad (7)$$

Therefore

$$\begin{aligned} \frac{\Phi_L}{4I} &= \frac{2\mu}{\mu + 1} \int_0^\pi \left(\frac{c}{R} \sin \theta + \frac{c^3}{R^3} \sin 3\theta + \dots \right) d\theta \\ &= \frac{4\mu}{\mu + 1} \left(\frac{c}{R} + \frac{c^3}{3R^3} + \frac{c^5}{5R^5} + \dots \right) \end{aligned} \quad (8)$$

If $c = R/2$

$$\frac{\Phi_L}{4I} = \frac{4\mu}{\mu + 1} \left(\frac{1}{2} + \frac{1}{24} + \frac{1}{160} + \dots \right)$$

Therefore

$$\frac{\Phi_L}{4I} \approx 2$$

For a two-turn coil the result will be

$$\frac{\Phi_L}{8I} \approx 2 \cos \alpha$$

Thus the leakage flux can be calculated very easily. Also, since it is independent of the permeability it will be negligibly small

comparison with the flux in the iron. Thus the inductance can be obtained either by an estimation of the flux crossing the diameter between the limits $r < c$, or $c < r < R$. Fig. 3 shows the remarkable way in which the surface polarity reduces the field in the region $r < c$ and increases it in the region $c < r < R$ by just that amount which makes the total flux very nearly the same for both these regions.

(4) EFFECT OF THE HOLES IN WHICH THE CONDUCTORS ARE PLACED

In the preceding discussion we have ignored any contribution to the magnetic field which might be made by the induced polarity on the walls of the holes in which the conductors are placed. It is now desired to take this effect into account.

(4.1) Effect of Small Circular Holes

Consider Fig. 8, which shows an iron cylinder of radius R which has a hole of radius b at its centre. A conductor carrying

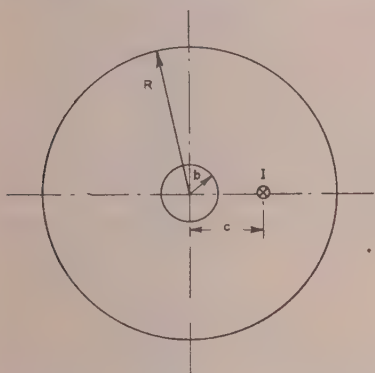


Fig. 8.—The effect of a small hole at the centre of the cylinder.

a current I is threaded through a hole of negligible dimensions at a distance c from the axis of the cylinder.

The induced surface pole strength on the surface of the cylinder has been estimated in Section 8.6.1. It is found to be given by the expression

$$2\pi\sigma_R = 2I \sum_{n=1}^{\infty} \frac{c^{2n} - b^{2n}}{R^{2n} - b^{2n}} \frac{R^{n-1}}{c^n} \sin n\theta \quad (9)$$

To estimate the effect of the hole, consider a cylinder for which $b = R/10$ and $c = R/2$. Then the fundamental in the Fourier series of the expression for the surface pole strength will be

$$2\pi\sigma_{R1} = 2I \frac{\left(4 - \frac{1}{100}\right)}{1 - \frac{1}{100}} \frac{2}{R} \sin \theta \simeq \frac{I}{R} \sin \theta$$

and this is also the expression if there is no hole.

Thus, to a very close approximation, the effect of the hole on the fundamental component of surface polarity at radius R is negligible. Consideration will show that the effect on the higher harmonics will be even less. This leads to the conclusion that the surface pole strength on any small hole within the cylinder does not appreciably affect the pole strength on the outside surface of the cylinder.

It is now proposed to calculate the field of the single-turn coil in Fig. 9, where the conductors are placed in holes of radius b . We shall make use of the above investigation to state that the

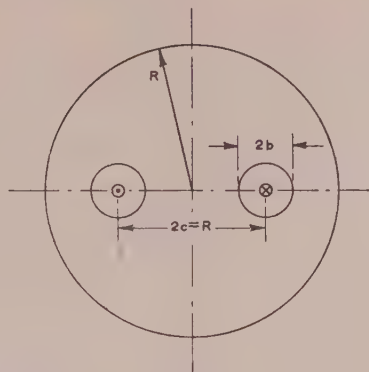


Fig. 9.—A single-turn coil threaded through holes in an iron cylinder.

polarity on each hole will not affect the polarity on the surface of the cylinder or the polarity on the other hole. Since the conductors are concentric with the holes, the polarity on each hole will thus be due only to the current in the other hole and the surface polarity of the cylinder. It is shown in Section 8.6 that the surface polarity on the holes will be given by

$$2\pi\sigma_b = 2I \left(\frac{2c}{R^2} \sin \phi - \frac{1}{2c} \sin \phi - \frac{b}{4c^2} \sin^2 \phi - \frac{b^2}{8c^3} \sin 3\phi - \dots \right) \quad (10)$$

If $c = R/2$ and $b = R/10$

$$2\pi\sigma_b = 2I \left(\frac{1}{R} \sin \phi - \frac{1}{R} \sin \phi - \frac{1}{10R} \sin 2\phi - \frac{1}{100R} \sin 3\phi - \dots \right)$$

Hence

$$2\pi\sigma_b \simeq -\frac{2I}{10R} \sin 2\phi$$

Thus for a spacing of $c = R/2$ the surface polarity σ_b consists, to a close approximation, only of a small second-harmonic term which is proportional to the ratio of the radius of the hole to the radius of the cylinder.

In order to estimate the effect of the hole on the self-inductance of the single-turn coil it is necessary to find H_θ due to σ_b between the holes. It is shown in Section 8.6 that this is given by

$$H_\theta = \frac{Ib^4}{2c^2r^3} \quad (11)$$

where r is measured from the centre of the hole. This field strength is a maximum at the surface of the hole and it is then

$$H_\theta = \frac{2}{100R} \text{ oersteds/AT}$$

Comparison with Section 2.1 will show that the field of the pole strength on the surface of the cylinder is given approximately by

$$H_\theta = \frac{2}{10R} \text{ oersteds/AT}$$

Thus the maximum field of the hole is only one-tenth of the field of the surface polarity on the cylinder, and this in turn is only a fraction of the field of the current in the coil. When it is remembered that the field of the hole decreases with the cube of

the distance from its centre, it can be safely concluded that the effect of the hole on the field is negligible.

If it is required to estimate the effect of the holes on the inductance of multi-turn coils, the position is substantially the same, so long as the holes are small.

Consider the two-turn coil in Fig. 5. In addition to the effects that have been estimated for a single-turn coil, it will now be necessary to take into consideration the action of the two conductors carrying current in the same direction on one another. In this case there will be an additional fundamental component of surface polarity on each hole as well as further harmonics. But the field strength of the fundamental will die away as the square of the distance from the hole, while the harmonics will die away even more rapidly. Thus, subject to the condition that $c \sin \alpha > b$, the effect of the holes can again be neglected. It should be noted, however, that this condition is more stringent than the condition $c > b$. If the holes are so close together that they almost coalesce, the problem is altered and it is no longer possible to neglect the effect of the holes on one another. For the solution of such problems graphical methods may well be found more convenient than the analytical methods used in the paper.

(4.2) Effect of Large Holes and Slots

Consider the problem with reference to Fig. 10. Surface polarity will be induced on the walls of each slot by the current within the slot as well as by the rest of the field. We shall not attempt to calculate the surface polarity, but will discuss the

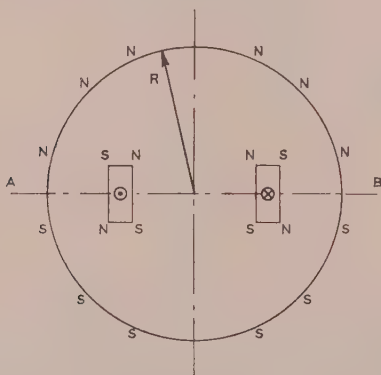


Fig. 10.—Surface polarity induced on the slots in which the conductors are placed.

problem in general terms only. It is clear that the surface polarity will contribute very appreciably to the field close to the slots and that this contribution will be such as to render the field strength across the diameter AB far more uniform than it would otherwise be. This can also be shown in a slightly different way as follows: The action of the slots will be to straighten out the field as indicated in Figs. 11 and 12. Consider a path like those shown dotted in the Figures. $\oint H dl$ will be zero in both cases, since no current is enclosed. In Fig. 12, where the field is parallel, there will be no contribution to $\oint H dl$ from the horizontal portions of the rectangular path shown. This means that H must be largely constant across the diameter. In Fig. 11, however, the field is not parallel, and hence H is not constant.

Therefore, whether the effect of the slot is considered from the point of view of surface polarity or of lines of magnetic force, the conclusion is the same. The field will be largely uniform

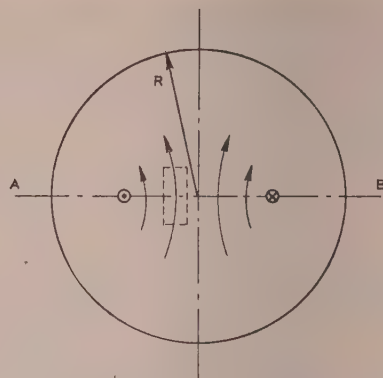


Fig. 11.—Magnetic field in the absence of slots.

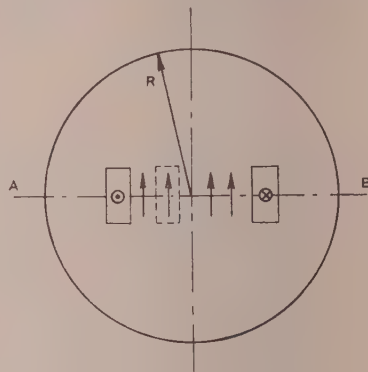


Fig. 12.—Magnetic field in the presence of slots.

between the conductors, and the self-inductance may be estimated by formulae that make use of this assumption.

If this is the conclusion, it may be asked whether the discussion in the paper is not a complete waste of time. For a transformer engineer, who places his windings in such large holes that they are properly called windows and not slots, there is indeed no need to worry about the non-uniformity of the flux distribution. But we have shown that the assumptions that are correct for the design of transformers are inadmissible in the design of rotating machines.

(5) APPROXIMATE FORMULAE FOR CALCULATION OF SELF-INDUCTANCE WHEN THE IRON IS UNSATURATED

The preceding Sections will be used in order to work out two special cases of coils threaded through holes in unsaturated iron cylinders.

(5.1) Self-Inductance of a Single-Turn Coil

The flux linkages in Fig. 9 will be those due to the current minus those due to the surface polarity on the cylinder. It is shown in Section 8.7.1 that the self-inductance is given by

$$L \simeq 4\mu \left[\log \left(\frac{R-b}{b} \right) - \frac{R-2b}{2R} - \frac{(R-2b)^3}{96R^3} - \frac{(R-2b)^5}{2560R^5} - \dots \right] \quad (1)$$

Hence

$$L \simeq 4\mu \left[\log \left(\frac{R-b}{b} \right) - \frac{R-2b}{2R} \right] \text{ electromagnetic units per unit length}$$

so long as μ does not vary greatly in the range of magnetic field strengths encountered. In any practical case it would be necessary to investigate the magnetic field close to the conductors to see whether the assumption was correct.

This formula for the self-inductance should be compared with the well-known formula for two parallel conductors $L' = 1 + 4 \log R/b$ electromagnetic units per unit length, where R is the distance between conductor centres and b is the conductor radius.

If this inductance is multiplied by the permeability of the iron it becomes

$$L' = \mu \left(1 + 4 \log \frac{R}{b} \right) \text{ electromagnetic units per unit length. } (13)$$

It is of interest to insert numerical values into the formulae. Let $b = R/10$.

$$\text{Then } L = 4\mu (\log 9 - \frac{2}{3}) \times 10^{-9} = 7.2\mu \times 10^{-9} \text{ henrys per unit length}$$

$$\text{and } L' = \mu(1 + 4 \log 10) \times 10^{-9} = 10.2\mu \times 10^{-9} \text{ henrys per unit length.}$$

Thus there is a considerable difference between the formulae, and it is essential to take the demagnetizing effect of the surface polarity into account.

(5.2) Self-Inductance of a Simple Phase-Advancer

The three-phase winding of a simple form of phase advancer is shown in Fig. 13. The conductors are threaded through small

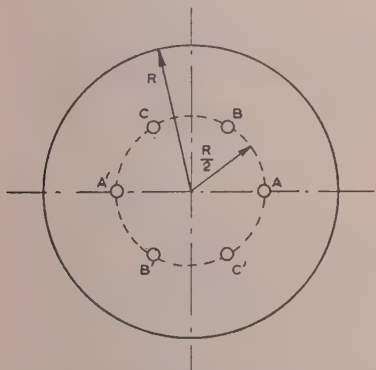


Fig. 13.—A simple form of phase advancer.

holes at a distance of $c = R/2$ from the axis of the cylinder. It is required to estimate the self-inductance per phase. It can be shown that this self-inductance is given by $L' = L + M$, where L is the self-inductance of each phase in the absence of the other phases,* The self-inductance L has already been found in Section 5.1 and is given by

$$L \simeq 4\mu \left[\log \left(\frac{R-b}{b} \right) - \frac{R-2b}{2R} \right] \text{ electromagnetic units per unit length}$$

The mutual inductance is calculated in Section 8.7.2. It is given by the expression

$$M = 8\mu \left\{ \left[\frac{R-2b}{R} \cos \alpha + \frac{1}{3} \left(\frac{R-2b}{R} \right)^3 \cos 3\alpha \right. \right.$$

$$\left. + \frac{1}{5} \left(\frac{R-2b}{R} \right)^5 \cos 5\alpha + \dots \right] - \left[\frac{R-2b}{4R} \cos \alpha + \frac{(R-2b)^3}{192R^3} \cos 3\alpha + \dots \right] \}$$

electromagnetic units per unit length (14)

where $\alpha = 60^\circ$.

If we take the case $b = R/10$

$$M \simeq 1.3\mu \text{ electromagnetic units per unit length}$$

Hence $L' = L + M = 8.5\mu \times 10^{-9}$ henry per phase per unit length. In a practical machine there are likely to be several turns per phase. It is not proposed to work out any further examples, but it will be seen that the method of this Section can readily be extended to cover such cases.

(6) CONCLUSIONS

It has been shown that magnetic theory makes it possible to obtain reliable estimates of the inductance of coils placed in holes in iron cylinders. It is found that a consideration of the surface polarity on the iron is essential to a solution of any particular case. This surface polarity can often be easily calculated and is found to be practically independent of the quality of the iron. It is thus possible to avoid having to make false assumptions about the flux distribution in the iron.

(7) ACKNOWLEDGMENT

The author gratefully acknowledges the help and encouragement given to him by Professor E. B. Moullin.

(8) APPENDICES

The mathematical tools used in the paper are extremely simple and have been in existence for a century. They are, however, not as well known as they deserve to be, and it may be of interest to give a description of them.

(8.1) Laplace's Equation in Two-Dimensional Polar Co-ordinates

The magnetic field of steady currents can be expressed in terms of a scalar potential function V , which has to satisfy Laplace's equation

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \theta^2} = 0 \quad . \quad . \quad . \quad (15)$$

This equation has the remarkable property that in its solution any term in r^n or $1/r^n$ has to be multiplied by $\cos n\theta$ or $\sin n\theta$. If this is realized explicitly, a great deal of labour can be saved in any particular solution.

(8.2) Boundary Conditions of the Magnetic Field at an Iron Surface

If iron is placed in a magnetic field and if this field has a component of magnetic force perpendicular to the surface of the iron, polarity will be induced on that surface. By the principle of superposition the resultant field can then be considered as the sum of the original inducing field and the field of the induced polarity.

Consider now a small area on the surface of the iron. Let the local surface polarity be denoted by σ and let the magnetic force on, and perpendicular to, the surface be denoted by H_p , where H_p is due to all magnets and currents whatever. Then

$$H_p + 2\pi\sigma = \mu(H_p - 2\pi\sigma)$$

* MOULLIN, E. B.: "Electromagnetic Principles of the Dynamo" (Oxford University Press, 1955).

Therefore

$$2\pi\sigma = \frac{\mu - 1}{\mu + 1} H_p \quad (16a)$$

In order to determine the surface polarity σ , it is thus necessary to find H_p . But, as already pointed out, H_p depends not only on the inducing field but also on the induced surface polarity. It appears that such problems can, in general, only be solved by successive approximations.

However, if the iron has a cylindrical surface, the problem is greatly simplified. It is shown below that the field of any distribution of induced polarity on a cylindrical surface has no perpendicular component anywhere on that surface. It follows that in such cases H_p is due to the inducing field only and is independent of σ . If the inducing field is known, the induced surface polarity can at once be found by the use of eqn. (16a).

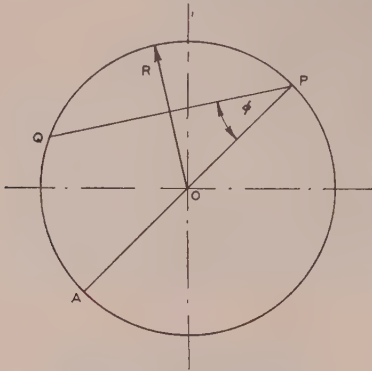


Fig. 14.—Shell of polarity.

This property of cylinders can be verified as follows. Fig. 14 depicts a cylindrical shell of polarity of radius R . Let the line density of polarity be denoted by σ . The normal force at P due to the filament of polarity at Q is

$$\delta H_p = \frac{2\sigma R \delta\theta}{QP} \cos \phi$$

Therefore

$$\delta H_p = \frac{2\sigma R \delta\theta}{AP} = \sigma \delta\theta$$

Hence the total perpendicular force at P due to the shell of polarity is

$$H_p = \int_0^{2\pi} \sigma d\theta$$

Since the total polarity is induced, it must be equal to zero, whence $H_p = 0$.

It has therefore been shown that, if an iron cylinder is placed in a magnetic field, the perpendicular component of magnetic force anywhere on the cylinder depends only on the inducing field and is independent of the induced polarity on the cylinder.

If the origin of co-ordinates is chosen on the axis of the cylinder, H_p (the perpendicular force) will be H_r (the radial force) and eqn. (16a) can be rewritten

$$2\pi\sigma = \frac{\mu - 1}{\mu + 1} H_r \quad (16b)$$

Consideration of this equation reveals that the surface polarity is almost independent of the permeability of the iron, since the ratio $(\mu - 1)/(\mu + 1)$ is always close to unity even for as low a value as $\mu = 100$. Thus the surface polarity is proportional to the inducing field and hence to the currents that give rise to this

field. The polarity is therefore largely independent of the magnetic quality of the iron, and the theory is not based on an incorrect assumption of constant permeability.

(8.3) Magnetic Field of a Current Filament I Distant c from the Origin of Co-ordinates

It is now proposed to determine the magnetic field of a very long current filament, which is not at the origin of co-ordinates. This problem is important because from its solution the field of any number of filaments can be obtained by superposition. The problem will be worked out in detail, in order to illustrate the use of Laplace's equation.

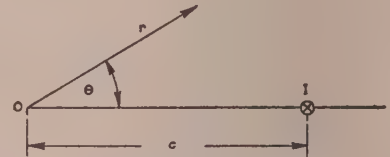


Fig. 15.—Current filament distant c from the origin.

Consider first in Fig. 15 the case $r < c$. At $\theta = 0$, it is known that

$$\frac{H_\theta}{2I} = \frac{1}{c - r} = \frac{1}{c} \left(1 - \frac{r}{c}\right)^{-1} = \frac{1}{c} + \frac{r}{c^2} + \frac{r^2}{c^3} + \dots$$

But
$$H_\theta = -\frac{1}{r} \frac{\partial V}{\partial \theta}$$

Therefore
$$-\frac{\partial V}{\partial \theta} \frac{1}{2I} = \frac{r}{c} + \frac{r^2}{c^2} + \dots$$

But r^n must be associated with $\sin n\theta$ or $\cos n\theta$. By symmetry $\sin n\theta$ is not admissible. Hence at a point r, θ ,

$$-\frac{\partial V}{\partial \theta} \frac{1}{2I} = \frac{r}{c} \cos \theta + \frac{r^2}{c^2} \cos 2\theta + \dots$$

Therefore
$$-\frac{V}{2I} = \sum_1^\infty \frac{r^n}{nc^n} \sin n\theta \quad (17)$$

and since $H_\theta = -\frac{1}{r} \frac{\partial V}{\partial \theta}$ and $H_r = -\frac{\partial V}{\partial r}$, the magnetic force

can be found at every point subject to the condition $r < c$.

Consider now the case $r > c$. At $\theta = 0$, it is known that

$$-\frac{H_\theta}{2I} = \frac{1}{r - c} = \frac{1}{r} \left(1 - \frac{c}{r}\right) = \frac{1}{r} + \frac{c}{r^2} + \frac{c^2}{r^3} + \dots$$

therefore
$$\frac{\partial V}{\partial \theta} \frac{1}{2I} = 1 + \frac{c}{r} + \frac{c^2}{r^2} + \dots$$

Hence at any point r, θ ($r > c$)

$$\frac{\partial V}{\partial \theta} \frac{1}{2I} = 1 + \frac{c}{r} \cos \theta + \frac{c^2}{r^2} \cos 2\theta + \dots$$

therefore
$$\frac{V}{2I} = \theta + \sum_1^\infty \frac{c^n}{nr^n} \sin n\theta \quad (18)$$

(8.4) Magnetic Field of a Distribution of Polarity of Form $\sigma_n \sin n\alpha$ on the Surface of an Iron Cylinder

Because the magnetic field of a current filament has been expressed in terms of a Fourier series, the induced polarity of the iron cylinder will also be of that form. In order to obtain the field of such a distribution, it becomes necessary to work out the problem in this Section.

Consider the field inside a cylinder of radius R . It can be shown by the method of Section 8.3 that the field of a filamentary polarity q at a distance R from the origin is given for $r < R$ by the expression

$$\frac{V}{2q} = \sum_1^{\infty} \frac{r^n}{nR^n} \cos n\theta$$

At $P(r, \theta)$ the magnetic field of a shell of polarity (Fig. 16) can therefore be described in terms of a potential function V ,

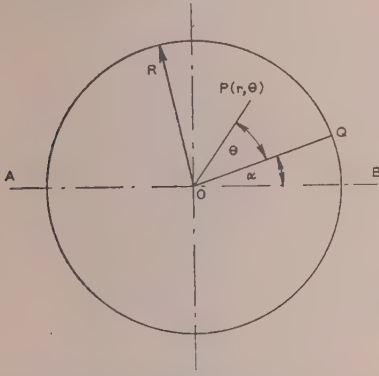


Fig. 16.—Cylindrical distribution of surface polarity.

which is the sum of the potentials due to filaments of polarity such as the typical one at Q .

Hence
$$\frac{\delta V}{2\sigma R \delta \alpha} = \sum_1^{\infty} \frac{r^n}{nR^n} \cos n(\theta - \alpha)$$

$\sigma = \sigma_n \sin n\alpha$, then

$$\frac{\delta V}{2\sigma_n R} = \sum_1^{\infty} \frac{r^n}{nR^n} (\cos n\theta \cos n\alpha + \sin n\theta \sin n\alpha) \sin n\alpha d\alpha$$

$$\frac{V}{2\sigma_n R} = \int_0^{2\pi} \left[\sum_1^{\infty} \frac{r^n}{nR^n} (\cos n\theta \cos n\alpha + \sin n\theta \sin n\alpha) \sin n\alpha d\alpha \right]$$

therefore
$$\frac{V}{2\sigma_n R} = \frac{\pi r^n}{nR^n} \sin n\theta$$

that
$$V = 2\pi\sigma_n \frac{r^n}{nR^{n-1}} \sin n\theta$$

ence, if the polarity is expressed as a Fourier series

$= \sum_1^{\infty} \sigma_n \sin n\alpha$, the potential will be given by the series

$$V = \sum_1^{\infty} 2\pi\sigma_n \frac{r^n}{nR^{n-1}} \sin n\theta \quad (19)$$

5. Magnetic Field of Certain Arrangements of Coils in the Presence of Iron Cylinders

In this Section it is proposed to derive the magnetic field for certain cases discussed in the paper.

5.1) Field of a Single-Turn Coil inside an Iron Cylinder.

By the use of eqn. (18) (see Figs. 1 and 2), the potential of two equal and oppositely directed current filaments is given in the region $r > c$ by

$$\frac{V}{2I} = \left(\theta + \sum_1^{\infty} \frac{c^n}{nr^n} \sin n\theta \right) - \left[\pi + \theta + \sum_1^{\infty} \frac{c^n}{nr^n} \sin n(\pi + \theta) \right]$$

Therefore
$$\frac{V}{4I} = -\frac{\pi}{2} + \sum_1^{\infty} \frac{c^n}{nr^n} \sin n\theta \quad (n \text{ odd}). \quad (20)$$

The surface polarity is given by the use of Section 8.2 by

$$2\pi\sigma = \frac{\mu - 1}{\mu + 1} H_r = -\frac{\mu - 1}{\mu + 1} \frac{\partial V}{\partial r} \simeq -\frac{\partial V}{\partial r} \quad (\text{at } r = R)$$

therefore
$$2\pi\sigma = 4I \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n}{R^{n+1}} \sin n\theta \quad (n \text{ odd}).$$

Hence the potential function of the field due to this surface polarity is, from Section 8.4,

$$\frac{V}{4I} = \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n\theta \quad (n \text{ odd}). \quad (21)$$

Consider now the magnetic field between the filaments. This will be due to the surface polarity and the current filaments. But the field due to the filaments can be obtained by the use of eqn. (17). It is given by

$$-\frac{V}{2I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta - \sum_1^{\infty} \frac{r^n}{nc^n} \sin n(\pi + \theta)$$

Therefore
$$-\frac{V}{4I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \quad (n \text{ odd}). \quad (22)$$

Hence the combined field of surface polarity and current filaments for $r < c$ will be given by

$$-\frac{V}{4I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta - \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n\theta \quad (n \text{ odd}). \quad (23)$$

Hence the field across the diameter for which $\theta = 0$ is given by

$$\frac{H_0}{4I} = -\frac{1}{r} \frac{\partial V}{\partial \theta} \frac{1}{4I} = \sum_1^{\infty} \frac{r^{n-1}}{c^n} - \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n r^{n-1}}{R^{2n}} \quad (n \text{ odd}).$$

Therefore

$$\frac{H_0}{4I} \simeq \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots \right) - \left(\frac{c}{R^2} + \frac{c^3 r^2}{R^6} + \frac{c^5 r^4}{R^{10}} + \dots \right). \quad (1)$$

Consider next the magnetic field in the region $c < r < R$. This is given by the sum of the potentials in eqns. (20) and (21). Hence

$$\frac{V}{4I} = -\frac{\pi}{2} + \sum_1^{\infty} \frac{c^n}{nr^n} \sin n\theta + \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n\theta \quad (n \text{ odd}).$$

Hence at $\theta = 0$, H_0 is given by

$$\frac{H_0}{4I} \simeq \left(\frac{c}{r^2} + \frac{c^3}{r^4} + \frac{c^5}{r^6} + \dots \right) + \left(\frac{c}{R^2} + \frac{c^3 r^2}{R^6} + \frac{c^5 r^4}{R^{10}} + \dots \right). \quad (5)$$

(8.5.2) Field of a Single-Turn Coil outside an Iron Cylinder.

The potential of the surface-polarity field in Fig. 4 can be obtained by the method of Section 8.5.1 and will be given by

$$\frac{V}{4I} = \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \quad (n \text{ odd}). \quad (24)$$

The potential of the field due to the current filaments is, as before,

$$-\frac{V}{4I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \quad (n \text{ odd}), \quad (22)$$

hence H_θ across the diameter $\theta = 0$ is given by

$$\frac{H_\theta}{4I} = \left(1 - \frac{\mu - 1}{\mu + 1}\right) \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right)$$

and
$$\frac{B_\theta}{4I} = \frac{\mu H_\theta}{4I} \simeq 2 \left(\frac{1}{c} + \frac{r^2}{c^3} + \frac{r^4}{c^5} + \dots\right) \quad (3)$$
 if $\mu \gg 1$.

(8.5.3) Field of a Two-Turn Coil inside an Iron Cylinder.

The field in Fig. 5 can be obtained from Section 8.5.1 by the superposition of the fields of two single-turn coils.

Thus for $r < c$

$$-\frac{V}{4I} = \left[\sum_1^{\infty} \frac{r^n}{nc^n} \sin n(\theta + \alpha) + \sum_1^{\infty} \frac{r^n}{nc^n} \sin n(\theta - \alpha) \right] - \frac{\mu - 1}{\mu + 1} \left[\sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n(\theta + \alpha) + \sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n(\theta - \alpha) \right]$$

Therefore

$$-\frac{V}{8I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \cos n\alpha - \frac{\mu - 1}{\mu + 1} \sum_1^{\infty} \frac{c^n r^n}{nR^{2n}} \sin n\theta \cos n\alpha \quad (25)$$

Hence for $\theta = 0$, H_θ is given by

$$\frac{H_\theta}{8I} \simeq \left(\frac{\cos \alpha}{c} + \frac{r^2 \cos 3\alpha}{c^3} + \frac{r^4 \cos 5\alpha}{c^5} + \dots \right) - \left(\frac{c \cos \alpha}{R^2} + \frac{c^3 r^2 \cos 3\alpha}{R^6} + \dots \right) \quad (4)$$

Similarly for $c < r < R$,

$$-\frac{H_\theta}{8I} \simeq \left(\frac{c \cos \alpha}{r^2} + \frac{c^3 \cos 3\alpha}{r^4} + \dots \right) + \left(\frac{c \cos \alpha}{R^2} + \frac{c^3 r^2 \cos 3\alpha}{R^6} + \dots \right) \quad (6)$$

(8.5.4) Expression for the Flux leaving the Iron Cylinder.

The radial magnetic force just outside the cylinder is given by

$$H_r + 2\pi\sigma = H_r \left(1 + \frac{\mu - 1}{\mu + 1}\right) = \frac{2\mu}{\mu + 1} H_r$$

Hence the flux leaving the cylinder is given by

$$\Phi_L = \frac{2\mu}{\mu + 1} \int_0^\pi H_r R d\theta \quad (7)$$

(8.6) Surface Polarity on Small Circular Holes

It is required in Fig. 8 to find the surface pole strength on the walls of the hole of radius b . Let this pole strength be denoted by σ_b and let the pole strength on the surface of the cylinder be denoted by σ_R . The magnetic field of the current filament has been calculated in Section 8.3 and is given by

$$-\frac{V}{2I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \quad \text{for } r < c \quad (17)$$

and
$$\frac{V}{2I} = \theta + \sum_1^{\infty} \frac{c^n}{nr^n} \sin n\theta \quad \text{for } r > c \quad (18)$$

The field of the polarity σ_R has been calculated in Section 8.4 and is given by

$$V = \sum_1^{\infty} 2\pi\sigma_R \frac{r^n}{nR^{n-1}} \sin n\theta \quad \text{for } r < R \quad (19)$$

By the method of Section 8.4 the field of the polarity σ_b can be shown to be

$$-V = \sum_1^{\infty} 2\pi\sigma_b \frac{b^{n+1}}{nr^n} \sin n\theta \quad \text{for } r > b \quad (26)$$

The boundary conditions give

at $r = R$, $2\pi\sigma_R = \frac{\mu - 1}{\mu + 1} \left(2I \frac{c^n}{R^{n+1}} - 2\pi\sigma_b \frac{b^{n+1}}{R^{n+1}} \right) \quad (27)$

at $r = b$, $2\pi\sigma_b = \frac{\mu - 1}{\mu + 1} \left(2I \frac{b^{n-1}}{c^n} - 2\pi\sigma_R \frac{b^{n-1}}{R^{n-1}} \right) \quad (28)$

If $\mu \gg 1$,
$$\frac{\mu - 1}{\mu + 1} \simeq 1$$

and
$$2\pi\sigma_R = 2I \frac{c^{2n} - b^{2n}}{R^{2n} - b^{2n}} \frac{R^{n-1}}{c^n} \quad (9)$$

It is required in Fig. 9 to find the surface polarity on the walls of the small holes of radius b . Let this pole strength be denoted by σ_b . For each hole it will be due to the polarity σ_R on the surface of the cylinder and also to the current I in the other hole. We shall neglect the effect of the holes on one another and shall regard σ_R as having only a fundamental ($\sin \theta$) component. Then from Section 8.5.1 the field of σ_R will be given by

$$\frac{V}{4I} = \frac{cr}{R^2} \sin \theta$$

This will give rise to a fundamental distribution of σ_b given by the expression

$$2\pi\sigma_b = 4I \frac{c}{R^2} \sin \phi \quad (29)$$

where ϕ is the angle measured with the centre of the hole as origin.

The field of the current is given in Section 8.3 as

$$-\frac{V}{2I} = \sum_1^{\infty} \frac{r^n}{nc^n} \sin n\theta \quad \text{for } r < c \quad (17)$$

It must therefore be written as

$$-\frac{V}{2I} = \sum_1^{\infty} \frac{r^n}{n(2c)^n} \sin n\phi \quad (30)$$

This gives rise to a pole strength σ_b given by

$$2\pi\sigma_b = -2I \sum_1^{\infty} \frac{b^{n-1}}{(2c)^n} \sin n\phi \quad (31)$$

Hence the total surface polarity σ_b is given by

$$2\pi\sigma_b = 2I \left(\frac{2c}{R^2} \sin \phi - \frac{1}{2c} \sin \phi - \frac{b}{4c^2} \sin 2\phi - \frac{b^2}{8c^3} \sin 3\phi - \dots \right) \quad (10)$$

The field of this polarity is given in eqn. (26). The tangential field of the second harmonic is therefore

$$H_\theta = -\frac{1}{r} \frac{\partial V}{\partial \theta} = \frac{Ib^4}{2c^2 r^3} \quad (11)$$

(8.7) Inductance Formulae

(8.7.1) Self-inductance of a Single-turn Coil.

The flux linkages in Fig. 9, due to the currents only, are given by

$$\Phi_1 = \mu \int_b^{2c-b} H_0 dx = 4I\mu \int_b^{2c-b} \frac{dx}{x} = 4I\mu \log \frac{2c-b}{b} \quad (32)$$

The flux linkages due to the surface polarity on the cylinder can be obtained from Section 8.5.1 by using the expression for H_0 in eqn. (1). If these flux linkages are denoted by Φ_2 ,

$$\Phi_2 = 2\mu \int_0^{c-b} H_0 dr = 8I\mu \left[\frac{cr}{R^2} + \frac{c^3 r^3}{3R^6} + \frac{c^5 r^5}{5R^{10}} + \dots \right]_0^{c-b} \\ \Phi_2 = I\mu \left[\frac{c(c-b)}{R^2} + \frac{c^3(c-b)^3}{3R^6} + \frac{c^5(c-b)^5}{5R^{10}} + \dots \right] \quad (33)$$

Hence the total flux linkages are given by

$$\Phi = \Phi_1 + \Phi_2 = 4I\mu \left[\log \frac{2c-b}{b} - \frac{2c(c-b)}{R^2} - \frac{2c^3(c-b)^3}{3R^6} - \dots \right] \quad (34)$$

Let $c = R/2$,

$$\Phi = 4I\mu \left[\log \frac{R-b}{b} - \frac{R-2b}{2R} - \frac{(R-2b)^3}{96R^3} - \frac{(R-2b)^5}{2560R^5} - \dots \right]$$

Hence the self-inductance is given by

$$\frac{\Phi}{I} = 4\mu \left[\log \frac{R-b}{b} - \frac{R-2b}{2R} - \frac{(R-2b)^3}{96R^3} - \frac{(R-2b)^5}{2560R^5} - \dots \right] \quad (12)$$

In this calculation the small amount of flux in the holes has been neglected and so has the effect of the polarity on the surface of the holes.

(8.7.2) Inductance per Phase of a Simple Phase Advancer.

Since we are considering in Fig. 13 a 3-phase machine, the inductance is given by the formula $L' = L + M$, where L is the self-inductance of one phase and M is the mutual inductance between any two phases. L is given by eqn. (12), and M can be found from eqn. (4).

$$M = 2\mu \int_0^{c-b} \frac{H_0}{2I} dr = 8\mu \int_0^{c-b} \left[\left(\frac{1}{c} \cos \alpha + \frac{r^2}{c^3} \cos 3\alpha + \dots \right) - \left(\frac{c}{R^2} \cos \alpha + \frac{c^3 r^2}{R^6} \cos 3\alpha + \dots \right) \right] dr \quad (35)$$

where $\alpha = 60^\circ$.

Hence

$$M = 8\mu \left\{ \left[\frac{c-b}{c} \cos \alpha + \frac{(c-b)^3}{3c^3} \cos 3\alpha + \dots \right] - \left[\frac{c(c-b)}{R^2} \cos \alpha + \frac{c^3(c-b)^3}{3R^6} \cos 3\alpha + \dots \right] \right\}$$

and if $c = R/2$

$$M = 8\mu \left\{ \left[\frac{R-2b}{R} \cos \alpha + \frac{(R-2b)^3}{3R^3} \cos 3\alpha + \dots \right] - \left[\frac{R-2b}{4R} \cos \alpha + \frac{(R-2b)^3}{192R^3} \cos 3\alpha + \dots \right] \right\} \quad (14)$$

AN APPROXIMATE METHOD FOR FINDING THE “BEST LINEAR SERVO MECHANISM”

By H. H. ROSENBROCK, B.Sc.(Eng.), Ph.D., Associate Member.

(The paper was first received 5th July, and in revised form 27th September, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

Wiener has given an explicit solution for the servo mechanism which minimizes the r.m.s. error when the system is following a signal contaminated with noise. The paper describes an elementary graphical method for evaluating Wiener's results, and makes use of a pair of transparent cursors which have already been described. An example is given, from which it appears that the accuracy of the method may be sufficient for engineering purposes.

An intermediate result of the graphical method gives the minimum phase corresponding to a given attenuation characteristic.

(1) INTRODUCTION

When a linear servo mechanism is following a signal containing noise, there will be an error due partly to the fact that the mechanism responds to the noise, and partly to the fact that it fails to respond perfectly to the signal. If the magnitude of this error is measured by its r.m.s. value, Wiener¹ has shown that there is a unique transfer function which makes the error of the system a minimum, and he gives an explicit solution for this function in terms of the frequency spectra of signal and noise.

In the following Section a graphical method for evaluating Wiener's solution is presented. It is based on a pair of transparent cursors which give respectively the Fourier sine and cosine transforms of a function. These cursors have already been described.²

(2) APPROXIMATE METHOD

The two transparent cursors which will be used give respectively

$$\frac{200}{\pi} \int_0^{\infty} \frac{F(\omega)}{\omega} \sin \omega t d\omega \quad \dots \quad (1)$$

and

$$\frac{200}{\pi} \int_0^{\infty} \frac{F(\omega)}{\omega} \cos \omega t d\omega \quad \dots \quad (2)$$

when they are applied to the function $F(\omega)$. We shall write these two quantities in the form

$$\mathcal{S}[F(\omega)] \quad \dots \quad (3)$$

and

$$\mathcal{C}[F(\omega)] \quad \dots \quad (4)$$

respectively, and shall refer to the corresponding cursors as the sine cursor and the cosine cursor. When the cursors are applied to $f(t)$ using t throughout the process in place of ω , and vice versa, there results

$$\mathcal{S}[f(t)] = \frac{200}{\pi} \int_0^{\infty} \frac{f(t)}{t} \sin \omega t dt \quad \dots \quad (5)$$

and

$$\mathcal{C}[f(t)] = \frac{200}{\pi} \int_0^{\infty} \frac{f(t)}{t} \cos \omega t dt \quad \dots \quad (6)$$

It will be assumed in what follows that the signal and noise are uncorrelated, and have respectively the power spectra $S(\omega)$ and $N(\omega)$. If there is correlation between signal and noise, their cross-spectral density must also be known, but no further difficulty is introduced into the solution. For the sake of generality it is assumed that the output of the servo mechanism is compared with the value which the signal takes α seconds later. If α is positive, the servo mechanism acts as a predictor, while if α is negative the servo mechanism is allowed to reproduce the signal with a delay of α seconds.

With these definitions it is shown in Section 9 that the closed-loop harmonic response $H(\omega)$ of the required servo mechanism is given by the following scheme:

$A(\omega)$ is the slope of $S(\omega) + N(\omega)$ in decibels per octave

$$B(t) = \frac{1}{100} \mathcal{S}[A(\omega)] \quad \dots \quad (7)$$

$$C(\omega) = \frac{15}{100} \mathcal{S}[B(t)] \text{ degrees} \quad \dots \quad (8)$$

$$D(t) = \frac{1}{100} \mathcal{C} \left\{ \frac{\omega S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \cos [57 \cdot 3 \alpha \omega + \frac{1}{2} C(\omega)] \right\} \\ - \frac{1}{100} \mathcal{S} \left\{ \frac{\omega S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \sin [57 \cdot 3 \alpha \omega + \frac{1}{2} C(\omega)] \right\} \quad \dots \quad (9)$$

$$H(\omega) = \frac{\pi}{400} \frac{e^{-j0.5C(\omega)}}{\sqrt{[S(\omega) + N(\omega)]}} \left\{ \mathcal{C}[tD(t)] - j \mathcal{S}[tD(t)] \right\} \quad (10)$$

Having obtained $H(\omega)$ we may find the r.m.s. error $\bar{\epsilon}$ by plotting³

$$y_1 = |H(\omega)|^2 \quad \dots \quad (11)$$

against

$$x_1 = \int_0^{\omega} N(\omega) d\omega \quad \dots \quad (12)$$

and

$$y_2 = |e^{j\alpha\omega} - H(\omega)|^2 \quad \dots \quad (13)$$

against

$$x_2 = \int_0^{\omega} S(\omega) d\omega \quad \dots \quad (14)$$

The r.m.s. error is obtained from the areas under these two curves by the formula

$$\bar{\epsilon} = \sqrt{\left(\int_{\omega=0}^{\infty} y_1 dx_1 + \int_{\omega=0}^{\infty} y_2 dx_2 \right)} \quad \dots \quad (15)$$

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Rosenbrock is with Costain-John Brown Ltd.

may be useful to remark that if $s(\tau)$ and $n(\tau)$ are the auto-correlation coefficients of signal and noise respectively,

$$S(\omega) = \frac{1}{100} \mathcal{E}[\tau s(\tau)] \quad . \quad . \quad . \quad (16)$$

$$N(\omega) = \frac{1}{100} \mathcal{E}[\tau n(\tau)] \quad . \quad . \quad . \quad (17)$$

and

$$x_2 = \int_0^\omega S(\omega) d\omega = \frac{1}{100} \mathcal{S}[s(\tau)] \quad . \quad . \quad . \quad (18)$$

$$x_1 = \int_0^\omega N(\omega) d\omega = \frac{1}{100} \mathcal{S}[n(\tau)] \quad . \quad . \quad . \quad (19)$$

Finally, if the servo mechanism has a single loop with unity feedback, we may find the open-loop harmonic response $G(\omega)$ from

$$G(\omega) = \frac{H(\omega)}{1 - H(\omega)} \quad . \quad . \quad . \quad (20)$$

The value of $G(\omega)$ will generally be well defined in the neighbourhood of the critical point $(-1, 0)$, but less well defined where $H(\omega)$ is nearly 1.

(3) EXAMPLE

As an example we take the following values:

$$S(\omega) = \frac{1}{1 + \omega^2} \quad . \quad . \quad . \quad (21)$$

$$N(\omega) = 0.2 \quad . \quad . \quad . \quad (22)$$

$$\alpha = 0 \quad . \quad . \quad . \quad (23)$$

The solution given by Wiener is

$$H(\omega) = \frac{1}{\sqrt{0.2} + \sqrt{1.2}} \frac{1}{\sqrt{1.2} + j\omega\sqrt{0.2}} \quad . \quad (24)$$

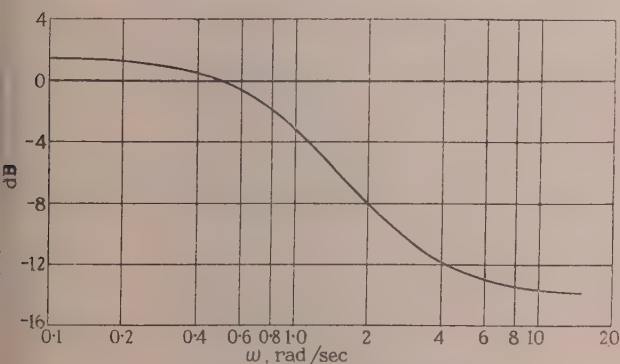


Fig. 1.—Graph of $20 \log_{10} [S(\omega) + N(\omega)]$.

Fig. 1 shows the graph of

$$20 \log_{10} [S(\omega) + N(\omega)] \quad . \quad . \quad . \quad (25)$$

The slope of this graph in decibels per octave is $A(\omega)$, and is shown in Fig. 2. The slope $A(\omega)$ was, for this example, obtained by calculation and not by measurement. For convenience the ordinates in Fig. 2 were multiplied by -0.4 before they were plotted. This factor is later removed by dividing the results obtained from the cursor by -0.4 .

On applying the sine cursor to Fig. 2 we obtain $B(t)$, which

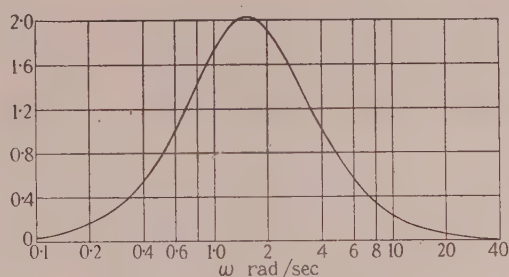


Fig. 2.—Graph of $-0.4A(\omega)$.

The graph gives the slope of the curve shown in Fig. 1.

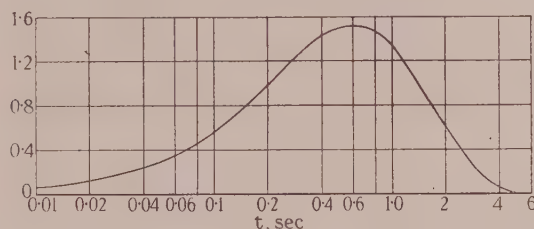


Fig. 3.—Graph of $-0.4B(t)$.

The graph was obtained from Fig. 2 by use of the sine cursor.

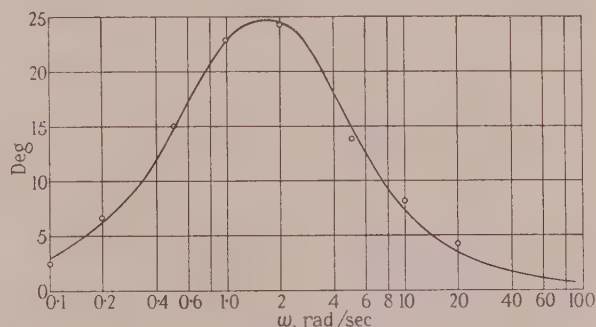


Fig. 4.—Graph of $-\frac{1}{2}C(\omega)$.

The graph was obtained from Fig. 3 by use of the sine cursor. The encircled points show calculated values for comparison. $C(\omega)$ is the minimum phase corresponding to the gain modulus shown in Fig. 1.

is shown in Fig. 3. A further application of the sine cursor gives $C(\omega)$, illustrated in Fig. 4. It will be noticed that if $A(\omega)$ has the constant value 6 dB per octave, then $B(t)$ will be 6 and $C(\omega)$ will be $15 \times 6 = 90^\circ$. This is what we should expect, since $C(\omega)$ is the phase of a minimum-phase system having $A(\omega)$ as the slope of its gain characteristic. The method for obtaining $C(\omega)$ from $A(\omega)$, given by eqns. (7) and (8), may be useful in other applications.

We now evaluate the quantities in eqn. (9) to which the cursors are to be applied; these are shown in Figs. 5 and 6. Performing the indicated operations we obtain $D(t)$, which when multiplied by t gives the result shown in Fig. 7. Some care is necessary at this point, since the value of $tD(t)$ obtained by the use of the cursor may be subject to error. For large values of t , $D(t)$ decreases exponentially to zero (either monotonically or in an oscillatory manner), but the error does not decrease proportionately. The value of $tD(t)$ will therefore fail to become zero and it will eventually increase. For this reason it is necessary to reject values of $D(t)$ obtained by means of the cursor unless they are several times larger than the likely error, which is about 2% of the greatest value assumed by $D(t)$.

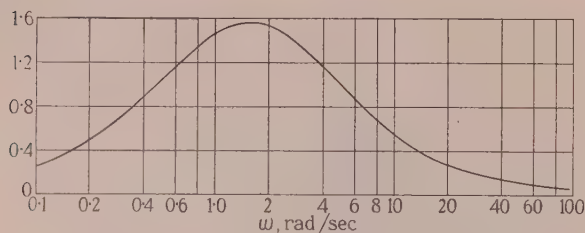


Fig. 5.—Graph of $-\frac{8}{3} \frac{\omega S(\omega) \cos \frac{1}{2}C(\omega)}{\sqrt{[S(\omega) + N(\omega)]}}$

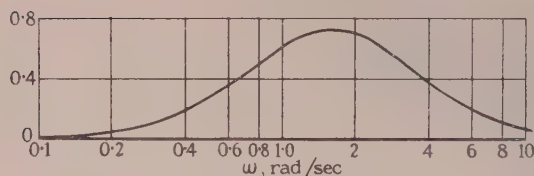


Fig. 6.—Graph of $\frac{8}{3} \frac{\omega S(\omega) \sin \frac{1}{2}C(\omega)}{\sqrt{[S(\omega) + N(\omega)]}}$

The difficulty mentioned in the last paragraph may be avoided by using, instead of eqn. (9),

$$-tD(t) = \frac{1}{100} \mathcal{L} \left(\omega \frac{d}{d\omega} \left\{ \frac{S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \cos [57 \cdot 3\alpha\omega + \frac{1}{2}C(\omega)] \right\} \right) + \frac{1}{100} \mathcal{L} \left(\omega \frac{d}{d\omega} \left\{ \frac{S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \sin [57 \cdot 3\alpha\omega + \frac{1}{2}C(\omega)] \right\} \right) \quad (26)$$

This formula, though less convenient than eqn. (9), will be preferable when there is any doubt about the behaviour of $tD(t)$ for large values of t .

For the example chosen it is evident from Figs. 5 and 6 and from the cursors that $tD(t)$ will behave in a simple manner as t becomes large. The last reliable point obtained from eqn. (9) is at $t = 3 \cdot 0$, and the curve is completed in a plausible manner, as shown in Fig. 7.

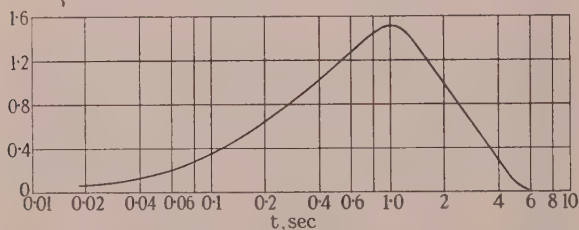


Fig. 7.—Graph of $3tD(t)$.

The graph was obtained from Figs. 5 and 6 by use of the cosine and sine cursors in accordance with eqn. (9).

The solution is completed by evaluating eqn. (10), and in Fig. 8 the results are compared with those calculated from (24). Fig. 9 shows the values of $G(\omega)$ obtained from eqn. (20).

The two integrals of eqn. (15) are evaluated from the graphs shown in Fig. 10. In the first graph, Fig. 10(a), we make use, in order to evaluate the area under the tail of the graph, of the fact that $|H(\omega)|$ is proportional to $1/\omega$ for large values of ω ; this follows from the angle at which $H(\omega)$ approaches the origin

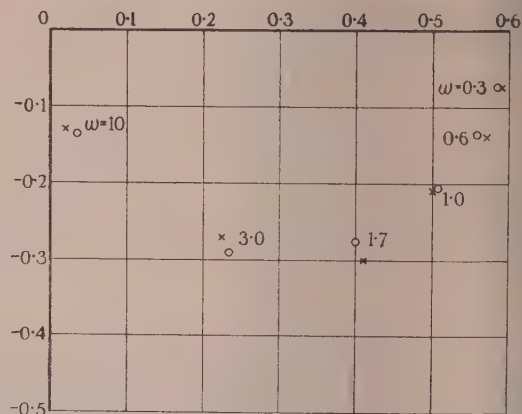


Fig. 8.—Points of the locus of $H(\omega)$.

The crosses show points obtained from Fig. 7 by means of the cosine and sine cursors in accordance with eqn. (10). The encircled points show calculated values for comparison.

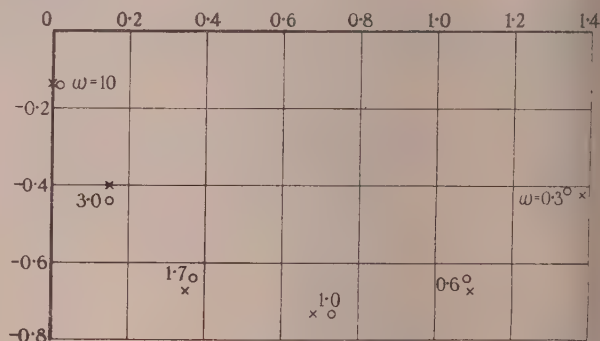


Fig. 9.—Points of the locus of $G(\omega)$.

The crosses show points obtained by means of the cursors. The encircled point shows calculated values for comparison.

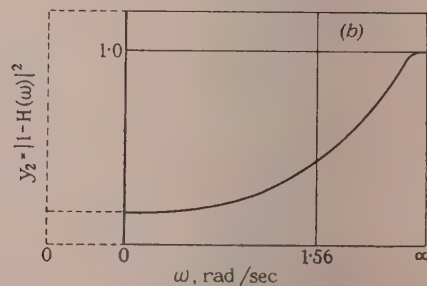
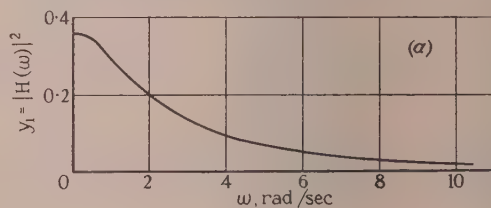


Fig. 10.—Graphs from which the r.m.s. error is obtained.

(a) The area under the graph gives the contribution to the mean square of error due to noise.
(b) The area under the graph gives the contribution to the mean square of error due to the failure of the system to respond perfectly to the signal.

The broken lines show the modification which must be made if the mean value of the signal is not zero.

and is shown in Section 4 to be a general result when $N(\omega)$ is a constant. We obtain

$$\bar{\epsilon} = \sqrt{(0.263 + 0.66)} = 0.96 \quad (27)$$

the calculated value being

$$\bar{\epsilon} = \sqrt{(0.2695 + 0.6412)} = 0.954 \quad (28)$$

It is evident from Figs. 8 and 9 that, although the accuracy with which individual points of $H(\omega)$ and $G(\omega)$ can be located is not high, the general behaviour of these functions is correctly shown. The value of $\bar{\epsilon}$ is likely to be obtained with good accuracy, because random errors in the individual points of $H(\omega)$ tend to cancel in eqn. (15).

(4) A SPECIAL CASE

The method presented in Section 2 can be simplified in the special case where the spectral density of the noise is independent of frequency, and $\alpha = 0$. It has been shown by Burt⁴ that under these conditions the required function $H(\omega)$ is given by

$$|1 - H(\omega)|^2 = \frac{N}{S(\omega) + N} \quad (29)$$

from this we readily find that

$$H(\omega) = 1 - \frac{\sqrt{N}}{\sqrt{[S(\omega) + N]}} e^{-j0.5C(\omega)} \quad (30)$$

where $C(\omega)$ is given by eqn. (8). Under the same conditions it can be shown that the least value of $\bar{\epsilon}$ is given by

$$\bar{\epsilon}^2 = 2N \int_0^\infty \mathcal{R}H(\omega) d\omega \quad (31)$$

where $H(\omega)$ is given by eqn. (30).

From eqn. (31) it can further be shown [at least when $S(\omega)$ is rational] that at high frequencies the required function $H(\omega)$ will be proportional to $1/\omega$ and will have a phase lag of 90° (see Section 9). This result is interesting, but the required behaviour can be achieved in practice only over a limited frequency range, since at high frequencies any practical system will have a phase lag greater than 90° . The frequency at which the phase lag of a practical system exceeds 90° can be increased by the use of lead networks, but only to a limited extent owing to the saturation which tends to result.

Eqns. (30) and (31) allow us in this special case to find $H(\omega)$ and the least value of $\bar{\epsilon}$ by two applications of the cursor and by drawing one graph. In the general case, six applications of the cursor are required to obtain $H(\omega)$, and two graphs must be drawn to find $\bar{\epsilon}$. It should be noted that eqn. (31) gives $\bar{\epsilon}$ only when $H(\omega)$ is given by eqn. (30); if we wish to find $\bar{\epsilon}$ for a practical system approximating to this $H(\omega)$ we must still draw two graphs and evaluate eqn. (15).

(5) DISCUSSION OF THE METHOD

In the design method presented by Wiener, $S(\omega)$ and $N(\omega)$ are approximated by rational functions, and these are dealt with analytically. The result is an analytical expression for $H(\omega)$ which is theoretically realizable by known techniques. The approximate method presented here results in a graphical formulation of $H(\omega)$, and there remains the problem of realizing the function by a physical system. A method by which this might be done has been presented elsewhere,² but it is likely that a different approach will be preferable for the following reasons.

The result given by the theory is "realizable" in the sense that

it could be obtained by a certain configuration of physical elements. This configuration is in itself, however, an idealization which neglects distributed parameters, stray capacitances, etc. Moreover the theory will demand a system which is characterized by a differential equation of a certain order. Quite apart from the stray effects already mentioned, it may happen that the important lumped parameters of the system are already represented by an equation of higher order than that required. It is then in theory possible to eliminate lags by appropriate lead networks, but in practice this might cause saturation of some elements and so be undesirable. An example of this difficulty has been given in the previous Section, where the theory demands a system having a phase lag of 90° at high frequencies.

A further difficulty is that the design method presented by Wiener cannot deal with signals which, because their mean value is different from zero, have a discontinuity in their power spectrum at $\omega = 0$. The reason for this is easy to see from Figs. 10(a) and 10(b). When the signal has a zero-frequency component, Fig. 10(b) must be modified by extending the scale of x_2 to the left of $\omega = 0$, as shown by the broken lines. The whole of the added portion represents the single frequency $\omega = 0$, and the curve of y_2 is extended across it with the value which it has for $\omega = 0$. The width of the added portion is such that the area above it, bounded by $y_2 = 1$, is equal to the power conveyed by the zero-frequency component. No modification has to be made to Fig. 10(a).

Since the conditions for $\omega > 0$ have not been changed in Figs. 10(a) and 10(b), the system given by Wiener's theory when the zero-frequency component is neglected will still result in the least possible contribution to the mean-square error from frequencies $\omega > 0$. There is now, however, a finite contribution to the mean-square error in Fig. 10(b) from the frequency $\omega = 0$, which we can eliminate only by making $|1 - H(\omega)| = 0$ for $\omega = 0$. Since an abrupt change of $H(\omega)$ at $\omega = 0$ is not physically realizable, $H(\omega)$ must be changed from the value given by Wiener's theory to $H(\omega) = 1$ in a small region near $\omega = 0$. This will result in an increase in the r.m.s. error due to frequencies $\omega > 0$, but on the other hand it will eliminate entirely the contribution from $\omega = 0$.

It can now be seen why the theory does not lead to a unique transfer function when the signal has energy at zero frequency. However quickly we make the transition to $H(\omega) = 1$, we can always find a system in which the transition is quicker. We can thus reduce the r.m.s. error as nearly as we like to the value which it would have if there were no zero-frequency component, but we can never attain this value. The difficulty is evidently a trivial one from the practical point of view, since we cannot in any case achieve more than an approximation to the required transfer function.

For these reasons it seems better to use the results obtained from the graphical method as a guide in modifying the system, rather than attempt to translate them directly into practice. If, for example, we are designing a linear system it may be important to know how much further improvement is possible. We may then evaluate Wiener's solution and find the corresponding r.m.s. error. If the system being designed gives an appreciably greater r.m.s. error, it may be modified using Wiener's solution as a guide. By plotting diagrams such as those in Fig. 10 for the system under consideration, it will be possible to see at what frequencies the contribution to the mean-square error most greatly exceeds its least possible value, and hence where improvement is most needed. If the signal has a zero-frequency component, this component should be neglected in applying the graphical method, and the locus of $H(\omega)$ for the system should follow that given by the method down to a low

value of ω , and should then tend to 1. In other words, for a single-loop system with unity feedback, $G(\omega)$ should follow Wiener's solution down to a low frequency* and should then tend to infinity.

The theory has been presented here in its application to servo mechanisms, but it applies equally to filters and predictors. The approximate method for evaluating Wiener's solution may be useful also in these fields.

(6) CONCLUSIONS

An approximate method has been given for evaluating Wiener's formula for the linear servo system having least r.m.s. error. The method, though tedious, is not impracticably so. It may serve as a guide in re-shaping the response locus of a system, and will indicate the least r.m.s. error which can be achieved.

(7) ACKNOWLEDGMENT

The author wishes to express his thanks to the Directors of Costain-John Brown Limited for permission to publish the paper.

(8) REFERENCES

- (1) WIENER, N.: "Extrapolation, Interpolation and Smoothing of Stationary Time Series" (Technology Press and John Wiley, New York, 1950).
- (2) ROSENBROCK, H. H.: "An Approximate Method for obtaining Transient Response from Frequency Response," *Proceedings I.E.E.*, Paper No. 1907 M, November, 1955 (102 B, p. 744).
- (3) ROSENBROCK, H. H.: "The Integral-of-Error-Squared Criterion for Servo Mechanisms," *ibid.*, Paper No. 1865 M, September, 1955 (102 B, p. 602).
- (4) BURT, E. G. C.: "Self-Optimizing Servo Systems," *Proceedings Cambridge Seminar on Non-Linear Control Problems*, September, 1954.
- (5) CAMPBELL, G. A., and FOSTER, R. M.: "Fourier Integrals for Practical Applications" (D. van Nostrand, New York, 1948), Nos. 445, 633, 448.1, 499.1.

(9) APPENDIX

The first stage in evaluating the required transfer function is to express $S(u) + N(u)$ as a product $\psi(u)\bar{\psi}(u)$, where $\psi(u + jv)$ is an analytic function having no singularities nor zeros for $v < 0$. Wiener¹ has shown that under suitable conditions

$$\psi(u + jv) = \exp \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\log [S(\omega) + N(\omega)]}{-j[\omega - (u + jv)]} d\omega \right\} \quad (32)$$

and that
$$\psi(u) = \lim_{v \rightarrow -0} \psi(u + jv) \quad (33)$$

Now if C is a contour consisting of the axis of u from $-R$ to $+R$ and a semicircle of radius R in the half-plane $v < 0$, we have formally

$$\int_C \left[\frac{d}{d\omega} \log \psi(\omega) \right] e^{-j\omega t} d\omega = 0 \quad (34)$$

We can easily show that the integral around the semicircle tends to zero as $R \rightarrow \infty$ for $t \geq \delta > 0$. Then, taking account of the

* This frequency must be above a certain minimum value if, in addition to the requirement for small r.m.s. error under working conditions, the system is required to complete its response to a step function within a given time.

odd and even symmetry of $\arg \psi(u)$ and $|\psi(u)|$ respectively we obtain

$$\int_{-\infty}^{\infty} \left[\cos ut \frac{d}{du} \arg \psi(u) - \sin ut \frac{d}{du} \log |\psi(u)| \right] du = 0 \quad (35)$$

whence for real values of ω

$$\begin{aligned} \int_0^{\infty} dt \cos \omega t \int_{-\infty}^{\infty} du \cos ut \frac{d}{du} \arg \psi(u) \\ = \int_0^{\infty} dt \cos \omega t \int_{-\infty}^{\infty} du \sin ut \frac{d}{du} \log |\psi(u)| \end{aligned} \quad (36)$$

This suggests that under suitable conditions

$$\frac{d}{d\omega} \arg \psi(\omega) = \frac{1}{\pi} \int_0^{\infty} dt \cos \omega t \int_{-\infty}^{\infty} du \sin ut \frac{d}{du} \log |\psi(u)| \quad (37)$$

whence, integrating from 0 to ω and noticing that

$$\log |\psi(u)| = \frac{1}{2} \log [S(u) + N(u)] \quad (38)$$

we have

$$\arg \psi(\omega) = \frac{1}{\pi} \int_0^{\infty} dt \frac{\sin \omega t}{t} \int_0^{\infty} du \sin ut \frac{d}{du} \log [S(u) + N(u)] \quad (39)$$

It may be mentioned here that a related formula is obtained by starting with $[\log \psi(u)]/\omega$ in place of $(d/d\omega) \log \psi(\omega)$ in eqn. (34). We thus obtain

$$\arg \psi(\omega) = \frac{\omega}{\pi} \int_0^{\infty} dt \cos \omega t \int_0^{\infty} du \sin ut \log \left[\frac{S(u) + N(u)}{S(0) + N(0)} \right] \quad (40)$$

This result may be used instead of eqn. (39), and it has the merit of avoiding the differentiation required in the latter. There are difficulties in applying the approximate method to eqn. (40), however, similar to those mentioned in connection with eqn. (10) of the paper.

We shall now verify eqn. (39) under the condition that $S(u) + N(u)$ is a rational function of u . This is not an unduly restrictive condition, since $S(u) + N(u)$ will be obtained by measurement, and we can always find a rational function of u which agrees with the measured values within the limits of experimental error. The function $S(u) + N(u)$ will be even and real, and we assume that it has no poles or zeros for real finite values of u .

Let
$$\psi(u) = \prod_{i=1}^n (ju + \alpha_i)^{r_i} \quad (41)$$

where
$$\psi(u)\bar{\psi}(u) = S(u) + N(u) \quad (42)$$

and
$$\alpha_i = \beta_i + j\gamma_i, \quad \beta_i > 0 \quad (43)$$

Then

$$\log \psi(u) = \sum_{i=1}^n r_i \left\{ \frac{1}{2} \log [(u + \gamma_i)^2 + \beta_i^2] + j \arctan \frac{u + \gamma_i}{\beta_i} \right\} \quad (44)$$

and
$$\frac{d}{du} \log |\psi(u)| = \sum_{i=1}^n r_i \left[\frac{u + \gamma_i}{(u + \gamma_i)^2 + \beta_i^2} \right] \quad (45)$$

Thus in eqn. (39) we have

$$\begin{aligned} du \sin ut \frac{d}{du} \log [S(u) + N(u)] \\ = \int_0^\infty du \sin ut \, 2 \sum_{i=1}^n r_i \left[\frac{u + \gamma_i}{(u + \gamma_i)^2 + \beta_i^2} \right] \end{aligned} \quad (46)$$

There are two types of integral to consider:

(a) If $\gamma = 0$, we have by known results⁵

$$2 \int_0^\infty du \sin ut \frac{u}{u^2 + \beta^2} = \pi e^{-\beta t}, \quad t > 0 \quad (47)$$

$$\frac{1}{\pi} \int_0^\infty dt \frac{\sin \omega t}{t} \pi e^{-\beta t} = \arctan \frac{\omega}{\beta} \quad (48)$$

(b) If $\gamma \neq 0$, then $-\gamma$ will be another value of γ_i and⁵

$$\begin{aligned} du \sin ut \left[\frac{u + \gamma}{(u + \gamma)^2 + \beta^2} + \frac{u - \gamma}{(u - \gamma)^2 + \beta^2} \right] \\ = 2 \int_0^\infty du \sin ut \left[\frac{u + j\beta}{(u + j\beta)^2 - \gamma^2} + \frac{u - j\beta}{(u - j\beta)^2 - \gamma^2} \right] \end{aligned} \quad (49)$$

$$= 2\pi \cos \gamma t e^{-\beta t}, \quad t > 0 \quad (50)$$

$$\int_0^\infty dt \frac{\sin \omega t}{t} 2\pi \cos \gamma t e^{-\beta t} = \arctan \frac{\omega + \gamma}{\beta} + \arctan \frac{\omega - \gamma}{\beta} \quad (51)$$

From eqns. (46), (48) and (51) we find that

$$\begin{aligned} \int_0^\infty dt \frac{\sin \omega t}{t} \int_0^\infty du \sin ut \frac{d}{du} \log [S(u) + N(u)] \\ = \sum_{i=0}^n r_i \arctan \frac{\omega + \gamma_i}{\beta_i} \end{aligned} \quad (52)$$

and this last expression, by eqn. (44), is $\arg \psi(\omega)$. This justifies eqn. (39) when $S(u) + N(u)$ is a rational function of u .

Now

$$\begin{aligned} \int_0^\infty du \sin ut \frac{d}{du} \log [S(u) + N(u)] \\ = \int_{u=0}^\infty \sin ut \frac{d}{d(\log u)} \log [S(u) + N(u)] d(\log u) \end{aligned} \quad (53)$$

$$= \frac{1}{20 \log_{10} e \log_e 2} \int_{u=0}^\infty \sin ut A(u) d(\log u) \quad (54)$$

$$= \frac{\pi}{2 \times 20 \log_{10} e \log_e 2} \frac{1}{100} \mathcal{S}[A(u)] \quad (55)$$

where $A(u)$ is the slope of $S(u) + N(u)$ in decibels per octave. It follows that

$$\begin{aligned} \arg \psi(\omega) &= 2 \arg \psi \\ &= \frac{2}{\pi} \int_{t=0}^\infty \sin \omega t \frac{\pi}{40 \log_{10} e \log_e 2} \frac{1}{100} \mathcal{S}[A(u)] d(\log t) \end{aligned} \quad (56)$$

or, in degrees,

$$C(\omega) = \frac{180}{\pi} \frac{\pi}{40 \log_{10} e \log_e 2} \frac{1}{100} \mathcal{S}[B(t)] \quad (57)$$

$$\simeq \frac{15}{100} \mathcal{S}[B(t)] \quad (58)$$

where

$$B(t) = \frac{1}{100} \mathcal{S}[A(u)] \quad (59)$$

We now have

$$\psi(u) = \sqrt{[S(u) + N(u)]} e^{j0.5C(u)} \quad (60)$$

and we have to evaluate Wiener's expression for $H(\omega)$:

$$H(\omega) = \frac{1}{2\pi\psi(\omega)} \int_0^\infty e^{-j\omega t} dt \int_{-\infty}^\infty \frac{S(u)e^{j\alpha u}}{\psi(u)} e^{jut} du \quad (61)$$

Now

$$\begin{aligned} \int_{-\infty}^\infty \frac{S(u)e^{j\alpha u}}{\psi(u)} e^{jut} du \\ = \int_{-\infty}^\infty \frac{S(u)}{\sqrt{[S(u) + N(u)]}} \left\{ \cos [\alpha u + \frac{1}{2}C(u)] \right. \\ \left. + j \sin [\alpha u + \frac{1}{2}C(u)] \right\} (\cos ut + j \sin ut) dt \end{aligned} \quad (62)$$

$$\begin{aligned} = 2 \int_0^\infty \frac{S(u)}{\sqrt{[S(u) + N(u)]}} \left\{ \cos [\alpha u + \frac{1}{2}C(u)] \cos ut \right. \\ \left. - \sin [\alpha u + \frac{1}{2}C(u)] \sin ut \right\} dt \end{aligned} \quad (63)$$

$$\begin{aligned} = \frac{\pi}{100} \mathcal{S} \left\{ \frac{S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \cos [57.3\alpha\omega + \frac{1}{2}C(\omega)] \right\} \\ - \frac{\pi}{100} \mathcal{S} \left\{ \frac{S(\omega)}{\sqrt{[S(\omega) + N(\omega)]}} \sin [57.3\alpha\omega + \frac{1}{2}C(\omega)] \right\} \end{aligned} \quad (64)$$

$$= \pi D(t) \quad (65)$$

where the angles in eqn. (64) are in degrees. Eqn. (10) of the paper then follows immediately.

To obtain eqn. (26) we integrate by parts the left-hand side of eqn. (62), obtaining

$$\pi D(t) = \frac{S(u)e^{j\alpha u}}{\psi(u)} \frac{e^{jut}}{jt} - \frac{1}{jt} \int_{-\infty}^\infty \frac{d}{du} \left[\frac{S(u)e^{j\alpha u}}{\psi(u)} \right] e^{jut} dt \quad (66)$$

The integrated part is zero for $t > 0$ since $S(u)$ tends to zero at either limit and $|\psi(u)| = \sqrt{[S(u) + N(u)]}$. Eqn. (26) then follows.

To obtain eqn. (31) we notice that under the conditions given in Section 4

$$\epsilon^2 = \int_0^\infty \{ H(\omega) \overline{H(\omega)} N + [1 - H(\omega)][1 - \overline{H(\omega)}] S(\omega) \} d\omega \quad (67)$$

$$\begin{aligned} = \int_0^\infty \{ [1 - H(\omega)][1 - \overline{H(\omega)}] [S(\omega) + N] \\ + [H(\omega) + \overline{H(\omega)}] N - N \} d\omega \end{aligned} \quad (68)$$

$$= N \int_0^\infty [H(\omega) + \overline{H(\omega)}] d\omega \quad (69)$$

by eqn. (29). Eqn. (31) follows immediately.

The value of $\bar{\epsilon}$ given by eqn. (69) will in all interesting cases be finite and not zero. Now $H(\omega)$ has no poles in the lower half-plane, and consequently $\int_C H(\omega) d\omega = 0$, where C is a contour consisting of the real axis from $-R$ to $+R$ and a semicircle of radius R in the lower half-plane. It follows that $\int_{-\infty}^{\infty} H(\omega) d\omega$ can only be finite and not zero if the integral of $H(\omega)$ round the large semicircle tends to a finite non-zero value.

If $H(\omega)$ is rational, this requires that it should behave like $1/j\omega$ at high frequencies. Finally

$$\int_{-\infty}^{\infty} H(\omega) d\omega = \int_{-\infty}^{\infty} \overline{H(\omega)} d\omega = \int_0^{\infty} [H(\omega) + \overline{H(\omega)}] d\omega$$

since each integral is real and $\Re H(\omega) = \Re \overline{H(\omega)}$. This shows that if $S(\omega)$ is rational [when $H(\omega)$ is also rational] and if the least value of $\bar{\epsilon}$ is finite and not zero, then at high frequencies $|H(\omega)|$ is proportional to $1/\omega$ and $\arg H(\omega) = -90^\circ$.

ANALYSIS OF LINEAR n -PORT NETWORKS

By I. CEDERBAUM.

(The paper was first received 24th August, and in revised form 15th October, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

Some properties of a general, non-reciprocal, linear network with n accessible terminal pairs (n -port) are investigated. With the use of a symbolic notation, a unique expression for all network functions is derived. This expression can be obtained in terms of the sub-determinants of the open-circuit impedance or the short-circuit admittance matrix of the n -port, and is applicable with any combination of open- or short-circuits at the network ports. Since all network functions, with such terminations of the n -port, may be expressed in terms of sub-determinants of one and the same matrix, numerous relations between them may be established. Some examples illustrate the application of this method.

(1) INTRODUCTION

The properties of linear networks with n accessible terminal pairs, or shortly linear n -ports, have been frequently discussed in the literature. Although the main interest has been concentrated on two-ports it has been observed that some problems, e.g. the theory of junctions in the microwave field,¹ the problems of conference telephony² and similar topics, may be profitably treated by the analysis of general n -ports. Recently some classical results of the theory of feedback amplifiers have been translated into the language of three-ports.³

Another interesting generalization of the idea of a network with two terminal pairs is a network with a set of n input and n output terminals leading to a multi-terminal transducer^{4,5,6} (m.t.t.).*

The successful solution of the synthesis problem of n -port networks obeying the reciprocity law^{7,8,9} has given a finishing touch to this branch of network theory. First attempts to deal with the synthesis of those linear n -ports which do not obey the reciprocity relations have now started.^{10,11}

The paper is devoted to the analysis of general linear n -ports which may violate the reciprocity law. The properties of an n -port may be described by any set of n independent equations between the $2n$ variables involved, i.e. the n currents and n voltages at the network ports. Of course, once such a system is given, all others may be obtained by solving the given equations for the specified unknowns. Any such system is represented by a matrix of its coefficients and is best adapted for some particular termination of the network. The short-circuit admittance matrix, for example, is particularly suitable if all ports are either terminated by voltage generators or short-circuited, whereas the open-circuit impedance matrix corresponds to the case of current generators or open-circuits at all ports of the network. There are, however, cases in practice where many terminations of an n -port are possible, and it may be necessary to deal with a considerable number of different representations. In such cases, the task of evaluating the numerous coefficients may prove extremely laborious.^{12†}

* It may help to grasp the difference between an n -port and an m.t.t. if the position of the hyphen is noted. The multi terminal-pair corresponds to the n -port and the multi-terminal pair to the m.t.t.

† For the two-port there are $(\frac{3}{2}) = 6$ possible systems with 24 coefficients which, for a network obeying the reciprocity law, are all given in Reference 12. The number of possible coefficients for a general n -port, K , is $(2n)!/[(n-1)!]^2$, and it may be extremely large, e.g. for $n = 4$, $K = 1120$.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Cederbaum is on leave of absence from the Scientific Department, Ministry of Defence, Israel, and is now at the Imperial College of Science and Technology, London.

The purpose of the paper is to present a symbolic notation which may be helpful in this respect and to derive some theorems valid for general linear n -ports.

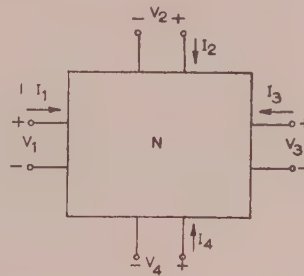


Fig. 1.—A four-port network.

(2) GENERAL SYSTEM OF EQUATIONS FOR AN n -PORT NETWORK

To fix attention let us consider a four-port network with the usual convention of voltage and current directions as illustrated in Fig. 1, and let its open-circuit representation be

$$\left. \begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 + Z_{13}I_3 + Z_{14}I_4 \\ V_2 &= Z_{21}I_1 + Z_{22}I_2 + Z_{23}I_3 + Z_{24}I_4 \\ V_3 &= Z_{31}I_1 + Z_{32}I_2 + Z_{33}I_3 + Z_{34}I_4 \\ V_4 &= Z_{41}I_1 + Z_{42}I_2 + Z_{43}I_3 + Z_{44}I_4 \end{aligned} \right\} \quad (1)$$

If voltage generators are applied, say, to the ports 1 and 3 and current generators to ports 2 and 4 the appropriate set of equations to be used is

$$\left. \begin{aligned} I_1 &= a_{11}V_1 + a_{13}V_3 + b_{12}I_2 + b_{14}I_4 \\ I_3 &= a_{31}V_1 + a_{33}V_3 + b_{32}I_2 + b_{34}I_4 \\ V_2 &= c_{21}V_1 + c_{23}V_3 + d_{22}I_2 + d_{24}I_4 \\ V_4 &= c_{41}V_1 + c_{43}V_3 + d_{42}I_2 + d_{44}I_4 \end{aligned} \right\} \quad (2)$$

The various coefficients a , b , c and d are evidently the network functions for specified terminations, e.g. b_{32} is the current ratio I_3/I_2 with ports 1 and 3 short-circuited and port 4 open-circuited, and d_{24} is the transfer impedance V_2/I_4 with port 2 open-circuited and ports 1 and 3 short-circuited.

To avoid confusion in the notation for minors of lower orders we shall denote the co-factor of the product of the elements located at the intersection of rows k , l , m , ... and columns r , s , t , ... respectively by $D_{klm...}^{rst...}$. Let D denote the determinant of the open-circuit impedance matrix of system (1). By solving the first and the third of eqns. (1) for I_1 and I_3 and introducing these values into the second and the fourth equation, it can be checked that system (2) can be rewritten in the form

$$\left. \begin{aligned} D_{24}^{24}I_1 &= D_{12}^{124}V_1 + D_{32}^{124}V_3 + D_{24}^{14}I_2 + D_{24}^{21}I_4 \\ D_{24}^{34}I_3 &= D_{12}^{324}V_1 + D_{32}^{324}V_3 + D_{24}^{34}I_2 + D_{24}^{31}I_4 \\ D_{24}^{24}V_2 &= -D_{14}^{24}V_1 - D_{34}^{24}V_3 + D_{24}^{42}I_2 - D_{24}^{41}I_4 \\ D_{24}^{24}V_4 &= -D_{14}^{24}V_1 - D_{34}^{24}V_3 - D_{24}^{42}I_2 + D_{24}^{41}I_4 \end{aligned} \right\} \quad (3)$$

The same results may be formally presented as a symbolic matrix equation as follows:

$$\begin{bmatrix} I_1 \\ I_3 \\ - \\ V_2 \\ V_4 \end{bmatrix} D_{24}^{24} = \begin{bmatrix} \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} & \begin{Bmatrix} 1 \\ 3 \end{Bmatrix} & \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} & \begin{Bmatrix} 1 \\ 4 \end{Bmatrix} \\ \begin{Bmatrix} 3 \\ 3 \end{Bmatrix} & \begin{Bmatrix} 3 \\ 3 \end{Bmatrix} & \begin{Bmatrix} 3 \\ 2 \end{Bmatrix} & \begin{Bmatrix} 3 \\ 4 \end{Bmatrix} \\ - & - & - & - \\ \begin{Bmatrix} 2 \\ 1 \end{Bmatrix} & \begin{Bmatrix} 2 \\ 3 \end{Bmatrix} & \begin{Bmatrix} 2 \\ 2 \end{Bmatrix} & \begin{Bmatrix} 2 \\ 4 \end{Bmatrix} \\ \begin{Bmatrix} 4 \\ 1 \end{Bmatrix} & \begin{Bmatrix} 4 \\ 3 \end{Bmatrix} & \begin{Bmatrix} 4 \\ 2 \end{Bmatrix} & \begin{Bmatrix} 4 \\ 4 \end{Bmatrix} \end{bmatrix} D_{24}^{24} \begin{bmatrix} V_1 \\ V_3 \\ - \\ I_2 \\ I_4 \end{bmatrix} \quad (4)$$

are negative, whereas the remaining elements have the positive sign. The co-factors standing on both sides are the principal minors with indices corresponding to known currents (or unknown voltages).

This form of representation is characteristic for those systems where at each port one of the variables—current or voltage—is fixed and the other unknown.* Let us take a general case of this type when the currents I_a, I_b, \dots, I_h and the voltages V_k, V_l, \dots, V_s are to be expressed in terms of the voltages V_a, V_b, \dots, V_h and the currents I_k, I_l, \dots, I_s , respectively, with $a, b, \dots, h, k, l, \dots, s$ representing some permutation of the indices $1, 2, 3, \dots, n$. In an analogous manner to eqn. (4) we have

$$D_{kl\dots s}^{kl\dots s} \begin{bmatrix} I_a \\ I_b \\ \vdots \\ I_h \\ - \\ V_k \\ V_l \\ \vdots \\ V_s \end{bmatrix} = \begin{bmatrix} \begin{Bmatrix} a \\ a \end{Bmatrix} & \begin{Bmatrix} a \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} a \\ h \end{Bmatrix} & \begin{Bmatrix} a \\ k \end{Bmatrix} & \begin{Bmatrix} a \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} a \\ s \end{Bmatrix} \\ \begin{Bmatrix} b \\ a \end{Bmatrix} & \begin{Bmatrix} b \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} b \\ h \end{Bmatrix} & \begin{Bmatrix} b \\ k \end{Bmatrix} & \begin{Bmatrix} b \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} b \\ s \end{Bmatrix} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \begin{Bmatrix} h \\ a \end{Bmatrix} & \begin{Bmatrix} h \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} h \\ h \end{Bmatrix} & \begin{Bmatrix} h \\ k \end{Bmatrix} & \begin{Bmatrix} h \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} h \\ s \end{Bmatrix} \\ - & - & - & - & - & - & - & - \\ \begin{Bmatrix} k \\ a \end{Bmatrix} & \begin{Bmatrix} k \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} k \\ h \end{Bmatrix} & \begin{Bmatrix} k \\ k \end{Bmatrix} & \begin{Bmatrix} k \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} k \\ s \end{Bmatrix} \\ \begin{Bmatrix} l \\ a \end{Bmatrix} & \begin{Bmatrix} l \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} l \\ h \end{Bmatrix} & \begin{Bmatrix} l \\ k \end{Bmatrix} & \begin{Bmatrix} l \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} l \\ s \end{Bmatrix} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \begin{Bmatrix} s \\ a \end{Bmatrix} & \begin{Bmatrix} s \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} s \\ h \end{Bmatrix} & \begin{Bmatrix} s \\ k \end{Bmatrix} & \begin{Bmatrix} s \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} s \\ s \end{Bmatrix} \end{bmatrix} D_{kl\dots s}^{kl\dots s} \begin{bmatrix} V_a \\ V_b \\ \vdots \\ V_h \\ - \\ I_k \\ I_l \\ \vdots \\ I_s \end{bmatrix} \quad (5)$$

where the square matrix standing on the right is a matrix of index pairs, and a "product" (or operation) of an index pair by a co-factor obeys the simple reduction rules of eqn. (4a) (with proper attention paid to the sequence of indices).

Of course, if, instead of the open-circuit impedance matrix we start with the short-circuit admittance matrix, then, denoting its determinant by H , we obtain a dual expression as follows:

$$H_{kl\dots s}^{kl\dots s} \begin{bmatrix} V_a \\ V_b \\ \vdots \\ V_h \\ - \\ I_k \\ I_l \\ \vdots \\ I_s \end{bmatrix} = \begin{bmatrix} \begin{Bmatrix} a \\ a \end{Bmatrix} & \begin{Bmatrix} a \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} a \\ h \end{Bmatrix} & \begin{Bmatrix} a \\ k \end{Bmatrix} & \begin{Bmatrix} a \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} a \\ s \end{Bmatrix} \\ \begin{Bmatrix} b \\ a \end{Bmatrix} & \begin{Bmatrix} b \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} b \\ h \end{Bmatrix} & \begin{Bmatrix} b \\ k \end{Bmatrix} & \begin{Bmatrix} b \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} b \\ s \end{Bmatrix} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \begin{Bmatrix} h \\ a \end{Bmatrix} & \begin{Bmatrix} h \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} h \\ h \end{Bmatrix} & \begin{Bmatrix} h \\ k \end{Bmatrix} & \begin{Bmatrix} h \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} h \\ s \end{Bmatrix} \\ - & - & - & - & - & - & - & - \\ \begin{Bmatrix} k \\ a \end{Bmatrix} & \begin{Bmatrix} k \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} k \\ h \end{Bmatrix} & \begin{Bmatrix} k \\ k \end{Bmatrix} & \begin{Bmatrix} k \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} k \\ s \end{Bmatrix} \\ \begin{Bmatrix} l \\ a \end{Bmatrix} & \begin{Bmatrix} l \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} l \\ h \end{Bmatrix} & \begin{Bmatrix} l \\ k \end{Bmatrix} & \begin{Bmatrix} l \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} l \\ s \end{Bmatrix} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \begin{Bmatrix} s \\ a \end{Bmatrix} & \begin{Bmatrix} s \\ b \end{Bmatrix} & \dots & \begin{Bmatrix} s \\ h \end{Bmatrix} & \begin{Bmatrix} s \\ k \end{Bmatrix} & \begin{Bmatrix} s \\ l \end{Bmatrix} & \dots & \begin{Bmatrix} s \\ s \end{Bmatrix} \end{bmatrix} H_{kl\dots s}^{kl\dots s} \begin{bmatrix} I_a \\ I_b \\ \vdots \\ I_h \\ - \\ V_k \\ V_l \\ \vdots \\ V_s \end{bmatrix} \quad (5a)$$

$$\begin{aligned} \left\{ \begin{matrix} p \\ j \end{matrix} \right\} D_{klm\dots}^{rst\dots} &= D_{jklm\dots}^{p\,rst\dots} & \left\{ \begin{matrix} l \\ j \end{matrix} \right\} D_{klm\dots}^{rst\dots} &= D_{klm\dots}^{rst\dots} \\ \left\{ \begin{matrix} p \\ t \end{matrix} \right\} D_{klm\dots}^{rst\dots} &= D_{klm\dots}^{rs\,p\dots} & \left\{ \begin{matrix} l \\ s \end{matrix} \right\} D_{klm\dots}^{rst\dots} &= D_{klm\dots}^{rst\dots} \end{aligned} \quad (4a)$$

Since

$$D_{klm\dots}^{rst\dots} = -D_{kml\dots}^{rst\dots} \text{ we have further } \left\{ \begin{matrix} l \\ t \end{matrix} \right\} D_{klm\dots}^{rst\dots} = -D_{klm\dots}^{rst\dots}$$

As is evident from eqn. (4) the upper index in each element of the square matrix is common to the whole row and equals the index of the corresponding element of the vector on the left side, whereas the lower or column index is that of the corresponding element of the vector on the right side of the equality.

If we partition this matrix conformably with the partitioning of the vectors of the variables into currents and voltages as indicated in eqn. (4), the elements of the bottom left-hand submatrix

The partitioning indicated in eqn. (5) corresponds to different dimensions of the matrix elements. The elements of the left-hand top submatrix are transfer or driving-point admittances, and those of the right-hand bottom submatrix are transfer or driving-point impedances, whereas the elements of the right-hand top and left-hand bottom submatrices are the current and voltage ratios respectively.

Remembering that fixing the voltage or the current at some port is to be regarded as short- or open-circuiting this port respectively, we may derive from the matrix equation (5) a unique expression for all network functions with any combination of some ports open- and the remaining short-circuited. As is clear

* If systems of equations were permitted in which at some ports both current and voltage are fixed (in which case at some other ports both must be unknown) the reduction rules assume more complicated forms. We limit ourselves here only to those important cases where at each port one variable is fixed and the other unknown.

from eqn. (5) any network function ψ_ν/ϕ_μ from the port μ to the port ν may be found from the symbolic equation

$$D_{kl\ldots s}^{kl\ldots s} \frac{\psi_\nu}{\phi_\mu} = \begin{matrix} + \\ - \end{matrix} \left\{ \begin{matrix} \nu \\ \mu \end{matrix} \right\} D_{kl\ldots s}^{kl\ldots s} \quad (6)$$

where ϕ and ψ may be currents or voltages, the indices k, l, \dots, s correspond to the open-circuited ports, the minus sign should be chosen if ϕ and ψ are voltages and the "product" of the index pair and the co-factor on the right-hand side of eqn. (6) obeys the above defined reduction rules. Of course, μ or ν (or both) may appear among the indices k, l, \dots, s of the open-circuited ports, and this always happens if ϕ is a current and/or ψ is a voltage, respectively.

(3) THE INVARIANT KERNEL OF THE SYMBOLIC REPRESENTATION

Eqn. (6) is equivalent to six different formulae. Denoting the open-circuited ports by a superscript on the network function we obtain from eqn. (6) the following:

The transfer admittance from the port b to a ($b \neq a$) with the ports k, l, m, \dots, s open-circuited is

$$\left(\frac{I_a}{V_b} \right)^{klm\ldots s} = Y_{ab}^{klm\ldots s} = \frac{D_{abklm\ldots s}^{klm\ldots s}}{D_{klm\ldots s}^{klm\ldots s}} \quad (7)$$

The transfer impedance from the port l to k ($l \neq k$) with the ports k, m, \dots, s open-circuited is

$$\left(\frac{V_k}{I_l} \right)^{km\ldots s} = Z_{kl}^{km\ldots s} = - \frac{D_{lm\ldots s}^{klm\ldots s}}{D_{klm\ldots s}^{klm\ldots s}} \quad (8)$$

The current ratio from the port k to a ($k \neq a$) with the ports l, m, \dots, s open-circuited is

$$\left(\frac{I_a}{I_k} \right)^{lm\ldots s} = \tau_{ak}^{lm\ldots s} = \frac{D_{kl\ldots s}^{al\ldots s}}{D_{kl\ldots s}^{kl\ldots s}} \quad (9)$$

The voltage ratio from port a to k ($a \neq k$) with the ports l, m, \dots, s open-circuited is

$$\left(\frac{V_k}{V_a} \right)^{kl\ldots s} = \tau_{ka}^{kl\ldots s} = - \frac{D_{al\ldots s}^{kl\ldots s}}{D_{kl\ldots s}^{kl\ldots s}} \quad (10)$$

Eqns. (7) and (8) with both port indices equal cover also the remaining two cases of driving-point admittances and impedances respectively, but then the sign in eqn. (8) must be made positive. An important property of the symbolic representation is the invariance of the matrix of index pairs under the changes in network termination. Let us, for example, consider the relation between the vector (I_1, V_2, V_3) and the vector (V_1, I_2, I_3) of a multi-port with all the remaining ports either short- or open-circuited:

$$\begin{bmatrix} I_1 \\ V_2 \\ V_3 \end{bmatrix} D_{23\ldots}^{23\ldots} = \begin{bmatrix} \begin{matrix} 1 \\ 1 \end{matrix} & \begin{matrix} 1 \\ 2 \end{matrix} & \begin{matrix} 1 \\ 3 \end{matrix} \\ -\begin{matrix} 2 \\ 1 \end{matrix} & \begin{matrix} 2 \\ 2 \end{matrix} & \begin{matrix} 2 \\ 3 \end{matrix} \\ -\begin{matrix} 3 \\ 1 \end{matrix} & \begin{matrix} 3 \\ 2 \end{matrix} & \begin{matrix} 3 \\ 3 \end{matrix} \end{bmatrix} \begin{bmatrix} V_1 \\ I_2 \\ I_3 \end{bmatrix}$$

is evident that the matrix of index pairs isolates the kernel, which is invariant with respect to any changes of the short- and open-circuits at all the remaining ports. The change in termination affects only the co-factor D whose indices, except for 2 and 3, correspond always to the open-circuited ports.

To illustrate the application of this property let us investigate the change of the network functions if a load is placed across the terminals of some port. To fix ideas let the variable impedance

Z be placed across the port 5 of a five-port and let us express the short-circuit current I_1 and the open-circuit voltage V_2 in terms of the given voltage V_3 and current I_4 . If a known voltage V_5 is assumed across Z , we have

$$\begin{bmatrix} I_1 \\ V_2 \end{bmatrix} D_{24}^{24} = \begin{bmatrix} \begin{matrix} 1 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 4 \end{matrix} \\ -\begin{matrix} 2 \\ 3 \end{matrix} & \begin{matrix} 2 \\ 4 \end{matrix} \end{bmatrix} \begin{bmatrix} \begin{matrix} 1 \\ 5 \end{matrix} \\ -\begin{matrix} 2 \\ 5 \end{matrix} \end{bmatrix} D_{24}^{24} \begin{bmatrix} V_3 \\ I_4 \\ V_5 \end{bmatrix} \quad (11)$$

If, on the other hand, a known current I_5 is assumed to be flowing through Z we have

$$\begin{bmatrix} I_1 \\ V_2 \end{bmatrix} D_{245}^{245} = \begin{bmatrix} \begin{matrix} 1 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 4 \end{matrix} \\ -\begin{matrix} 2 \\ 3 \end{matrix} & \begin{matrix} 2 \\ 4 \end{matrix} \end{bmatrix} \begin{bmatrix} \begin{matrix} 1 \\ 5 \end{matrix} \\ \begin{matrix} 2 \\ 5 \end{matrix} \end{bmatrix} D_{245}^{245} \begin{bmatrix} V_3 \\ I_4 \\ I_5 \end{bmatrix} \quad (12)$$

We may notice the invariance of the two first columns of the rectangular matrix in eqns. (11) and (12) and the change in the indices of the co-factors. Let us partition the matrices on the right-hand side, as indicated by the lines. Eqns. (11) and (12) may then be rewritten as

$$\begin{bmatrix} I_1 \\ V_2 \end{bmatrix} D_{24}^{24} = \begin{bmatrix} \begin{matrix} 1 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 4 \end{matrix} \\ -\begin{matrix} 2 \\ 3 \end{matrix} & \begin{matrix} 2 \\ 4 \end{matrix} \end{bmatrix} \begin{bmatrix} V_3 \\ I_4 \end{bmatrix} + \begin{bmatrix} \begin{matrix} 1 \\ 5 \end{matrix} \\ -\begin{matrix} 2 \\ 5 \end{matrix} \end{bmatrix} D_{24}^{24} V_5 \quad (13)$$

$$\begin{bmatrix} I_1 \\ V_2 \end{bmatrix} D_{245}^{245} = \begin{bmatrix} \begin{matrix} 1 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 4 \end{matrix} \\ -\begin{matrix} 2 \\ 3 \end{matrix} & \begin{matrix} 2 \\ 4 \end{matrix} \end{bmatrix} \begin{bmatrix} V_3 \\ I_4 \end{bmatrix} + \begin{bmatrix} \begin{matrix} 1 \\ 5 \end{matrix} \\ \begin{matrix} 2 \\ 5 \end{matrix} \end{bmatrix} D_{245}^{245} I_5 \quad (14)$$

The coefficients of V_5 in eqn. (13) and I_5 in eqn. (14) are equal since

$$\begin{Bmatrix} 1 \\ 5 \end{Bmatrix} D_{24}^{24} = D_{245}^{241} = \begin{Bmatrix} 1 \\ 5 \end{Bmatrix} D_{245}^{245}$$

and

$$-\begin{Bmatrix} 2 \\ 5 \end{Bmatrix} D_{24}^{24} = -D_{245}^{254} = \begin{Bmatrix} 2 \\ 5 \end{Bmatrix} D_{245}^{245}$$

If now an impedance Z is placed across port 5, $V_5 = -ZI_5$. Multiplying eqn. (14) by Z and adding to eqn. (13) gives

$$\begin{bmatrix} I_1 \\ V_2 \end{bmatrix} (D_{24}^{24} + ZD_{245}^{245}) = \begin{bmatrix} \begin{matrix} 1 \\ 3 \end{matrix} & \begin{matrix} 1 \\ 4 \end{matrix} \\ -\begin{matrix} 2 \\ 3 \end{matrix} & \begin{matrix} 2 \\ 4 \end{matrix} \end{bmatrix} (D_{24}^{24} + ZD_{245}^{245}) \begin{bmatrix} V_3 \\ I_4 \end{bmatrix} \quad (15)$$

since the two last terms cancel out.

Eqn. (15) like eqn. (6) is equivalent to six relations giving explicitly the different network functions as functions of Z . Thus, for example, the short-circuit current ratio $I_1:I_4$ with port 2 open-circuited* is

$$\left(\frac{I_1}{I_4} \right)^2 = \tau_{14}^2(Z) = \frac{D_{24}^{21} + ZD_{245}^{215}}{D_{24}^{24} + ZD_{245}^{245}}$$

and similarly for other network functions, restating the known fact that each network function is a rational linear function of any network impedance.¹³

(4) SOME PROPERTIES OF GENERAL n -PORT NETWORKS

Since, as follows from eqns. (7)–(10), each network function for any combination of short- and open-circuits at the network

* It should be noticed that the superscripts 2 in the first two members of this equation do not indicate squares but open-circuiting the port 2.

ports may be expressed as a ratio of some two co-factors of one and the same determinant, many interesting relations between them may be established. Some illustrative examples are given as follows:

Let us, for example, consider a three-port. Using eqns. (7)–(10) we see that

$$\frac{Y_{12}Y_{23}Y_{31}}{Y_{21}Y_{32}Y_{13}} = \frac{Z_{12}^2Z_{23}^2Z_{31}^2}{Z_{21}^2Z_{32}^2Z_{13}^2} = \frac{\tau_{12}\tau_{23}\tau_{31}}{\tau_{21}\tau_{32}\tau_{13}} = \frac{t_{12}^2t_{23}^2t_{31}^2}{t_{21}^2t_{32}^2t_{13}^2} = \frac{D_2^2D_3^2D_1^2}{D_1^2D_2^2D_3^2} \quad (16)$$

which may be expressed as follows:

If the ports of the network are arranged in some order, the ratio of the products of the transfer functions (τ , t) between two neighbouring ports taken in one direction and its reverse is equal to the ratio of the products of transfer admittances or transfer impedances taken in the same way, proper attention being paid to the port terminations.

For a general two-port and for an n -port obeying the reciprocity law, eqn. (16) is trivial, and the above ratio is equal to unity.

The influence of opening a port on the transfer admittance or transfer impedance between two other ports in a general linear network is reflected in the following relations:

$$Y_{12} - Y_{12}^3 = \frac{D_2}{D} - \frac{D_{23}^3}{D_3} = \frac{D_2^3D_3 - DD_{23}^3}{DD_3} = \frac{D_2^3D_3}{DD_3} = \tau_{13}Y_{32} \quad (17)$$

$$\text{and} \quad Z_{12}^{13} - Z_{12}^1 = -\frac{D_{23}^{13}}{D_{123}^{13}} + \frac{D_2^1}{D_{12}^1}$$

Since all co-factors appearing here may be derived from the common co-factor D_2^1 , if we denote D_2^1 by F we have

$$Z_{12}^{13} - Z_{12}^1 = \frac{F_3^3}{F_{13}^{23}} - \frac{F}{F_1^2} = \frac{F_3^3F_1^2 - FF_{13}^{23}}{F_{13}^{23}F_1^2} = \frac{F_1^3F_3^2}{F_1^2F_{13}^{23}} = \frac{D_{21}^{13}D_{23}^{12}}{D_{12}^{12}D_{13}^{23}} = t_{13}^2Z_{32}^{13} \quad (18)$$

Eqn. (17), for example, shows that the change in the transfer admittance from generator 2 to receiver 1 caused by open-circuiting switch 3 is proportional both to the short-circuit transfer admittance from the generator to the switch and to the short-circuit current ratio from the switch to the receiver.

Equating the indices 1 and 2 in eqn. (17) and afterwards replacing the index 3 by 2, we get the influence of opening a port on the driving-point admittance:

$$Y_{11} - Y_{11}^2 = \tau_{12}Y_{21}$$

$$\text{or} \quad 1 - \frac{Y_{11}^2}{Y_{11}} = \tau_{12}\tau_{21} = t_{12}^2t_{21}^2 \quad (19)$$

The right-hand side of eqn. (19) is apparently not influenced by the change between the indices 1 and 2. This expresses the fact that the relative change in the driving-point admittance is the same if the point of observation and the switch change their places, and it is equal to the product of the current or voltage ratios in both directions.

The influence of open-circuiting a port on the short-circuit current ratio between two other ports may be obtained from

$$\tau_{12} - \tau_{12}^3 = \frac{D_2}{D_2^2} - \frac{D_{23}^{13}}{D_{23}^{23}} = \frac{D_2^1D_{23}^{23} - D_2^2D_{23}^{13}}{D_2^2D_{23}^{23}}$$

Using the identity¹⁴

$$D_{23}^{23}D_2^1 + D_{23}^{31}D_2^2 + D_{23}^{12}D_2^3 = 0$$

$$\text{we obtain} \quad \tau_{12} - \tau_{12}^3 = \frac{D_{32}^{12}D_2^3}{D_{23}^{23}D_2^2} = \tau_{13}^2\tau_{32} \quad (20)$$

Similarly the change in the open-circuit voltage ratio becomes

$$t_{12}^{13} - t_{12}^1 = -\frac{D_{23}^{13}}{D_{13}^{23}} + \frac{D_2^1}{D_1^2} = \frac{D_{13}^{13}D_2^1 + D_{32}^{13}D_1^2}{D_{13}^{23}D_1^2} = \frac{D_{12}^{13}D_3^1}{D_{13}^{23}D_1^1} = t_{13}^2t_{32}^{13} \quad (21)$$

Example.—Let us find the transfer admittance

$$Y_{21} = \frac{I_2}{V_1}$$

of the feedback amplifier illustrated in Fig. 2 when the switch 3 is short-circuited.

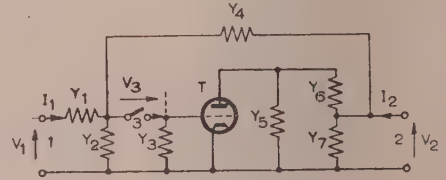


Fig. 2.—Feedback amplifier (with switch 3 short-circuited).

Let us denote the transconductance of the valve T by g , and let its anode admittance be incorporated in Y_5 . Using eqn. (17) (with the indices 1 and 2 interchanged) we have

$$Y_{21} = Y_{21}^3 + \tau_{23}Y_{31} \quad (22)$$

All the network functions standing on the right-hand side may be easily evaluated.

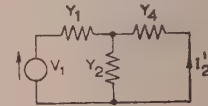


Fig. 3.—Circuit of Fig. 2 with switch 3 open-circuited and port 2 short-circuited.

The transfer admittance Y_{21}^3 with the switch 3 open-circuited may be found from the passive network of Fig. 3 as

$$Y_{21}^3 = \frac{I_2'}{V_1} = -\frac{Y_1Y_4}{Y_1 + Y_2 + Y_4} \quad (23)$$

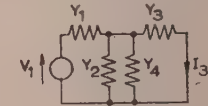


Fig. 4.—Circuit of Fig. 2 with ports 2 and 3 short-circuited.

The transfer admittance Y_{31} with the output 2 short-circuited may be evaluated from the passive network of Fig. 4, giving

$$Y_{31} = \frac{I_3}{V_1} = \frac{Y_1Y_3}{Y_1 + Y_2 + Y_3 + Y_4} \quad (24)$$

The current ratio τ_{23} with the input short-circuited may be obtained from the straightforward amplifier circuit of Fig. 5. We have

$$I_2'' = I_2'' + I_2^{iv}$$

where

$$\frac{I_2''}{I_3} = \frac{Y_4}{Y_1 + Y_2 + Y_4} \quad (25)$$

$$\frac{I_2^{iv}}{I_3} = \frac{g}{Y_3} \frac{Y_6}{Y_5 + Y_6} \quad (26)$$

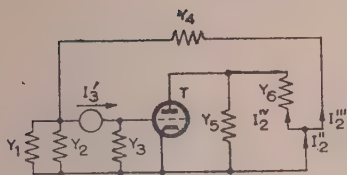


Fig. 5.—Circuit of Fig. 2 with port 2 short-circuited and port 3 fed by a current generator.

Thus from eqn. (22) we have for the desired transfer admittance

$$Y_{21} = -\frac{Y_1 Y_4}{Y_1 + Y_2 + Y_4} + \frac{Y_1 Y_3}{Y_1 + Y_2 + Y_3 + Y_4} \times \left[\frac{Y_4}{Y_1 + Y_2 + Y_4} + \frac{g Y_6}{Y_3(Y_5 + Y_6)} \right] \\ = \frac{Y_1(g Y_6 - Y_4 Y_5 - Y_4 Y_6)}{(Y_1 + Y_2 + Y_3 + Y_4)(Y_5 + Y_6)} \quad (27)$$

In conclusion we wish to point out that the examples have been chosen merely to illustrate the advantages offered by the described method. With the same techniques, other theorems concerning general linear n -ports may easily be produced.

(5) ACKNOWLEDGMENT

The author wishes to express his gratitude to Dr. A. Talbot, Imperial College of Science and Technology, London, for his continued interest in this work and to the Scientific Department, Ministry of Defence, Israel, under whose auspices the research was done.

(6) REFERENCES

- (1) MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: "Principles of Microwave Circuits, Radiation Laboratory Series No. 8" (McGraw-Hill Book Co. Inc., 1948), pp. 124–127.
- (2) BELEVITCH, V.: "Theory of $2n$ -Terminal Networks with Applications to Conference Telephony," *Electrical Communication*, 1950, 27, p. 231.
- (3) JOB, F.: "Sur quelques propriétés des réseaux linéaires a trois paires de bornes et leurs applications a la réalisation des impédances négatives et des amplificateurs a réaction négative," *Câbles et Transmission*, 1954, 8, p. 219.
- (4) RICE, S. O.: *Bell System Technical Journal*, 1941, 20, p. 131.
- (5) COLLET, L.: "Définition et caractéristiques de circuits superposés," *Annales des Télécommunications*, 1949, 4, p. 42.
- (6) SHEKEL, J.: "Matrix Analysis of Multi-Terminal Transducers," *Proceedings of the Institute of Radio Engineers*, 1954, 42, p. 840.
- (7) BAYARD, M.: "Synthèse des réseaux passifs à un nombre quelconque de paires de bornes connaissant leurs matrices d'impédance ou d'admittance," *Bulletin de la Société Française des Electriciens*, 1949, 2, p. 497.
- (8) OONO, Y.: "Synthesis of a Finite $2n$ -Terminal Network by a Group of Networks each of which contains only One Ohmic Resistance," *Journal of Mathematics and Physics*, 1950, 29, p. 13.
- (9) BELEVITCH, V.: "Synthèse des réseaux électriques passifs à n paires de bornes de matrice de répartition prédéterminée," *Annales des Télécommunications*, 1951, 6, p. 302.
- (10) OONO, Y., and YASUURA, K.: "Synthèse des réseaux passifs a n paires de bornes donnés par leurs matrices de répartition," *ibid.*, 1954, 9, p. 73.
- (11) CARLIN, H. J.: "On the Physical Realizability of Linear Non-Reciprocal Networks," *Proceedings of the Institute of Radio Engineers*, 1955, 43, p. 608.
- (12) GUILLEMIN, E. A.: "Communication Networks," Vol. II (John Wiley, 1935), pp. 137–138.
- (13) BODE, H. W.: "Networks Analysis and Feedback Amplifier Design" (Van Nostrand, 1945), p. 10.
- (14) AITKEN, A. C.: "Determinants and Matrices" (Oliver and Boyd, 1951), p. 103.

THE EDDY-CURRENT ANOMALY IN ELECTRICAL SHEET STEEL

By H. ASPDEN, Ph.D., B.Sc., Wh.Sc., Graduate.

(The paper was first received 14th September, and in revised form 7th November, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

A theory which accounts for the well-known discrepancy between predicted eddy-current losses in electrical sheet steels and the experimentally observed values is presented. The anomaly is shown to be due partly to the magnetic inhomogeneity arising from ferromagnetic domain structure and partly to a time-lag effect caused by the finite domain boundary movements. A new experimental approach speed of to the study of the eddy-current anomaly is described. This involves the use of a method of measuring the anomaly factor as it applies instantaneously at a point in a magnetization cycle.

LIST OF PRINCIPAL SYMBOLS

- b = Half-thickness of electrical sheet, cm.
 n = Number of turns of magnetizing coil per centimetre length.
 ρ = Resistivity of electrical sheet, ohm-cm.
 t = Time, sec.
 ω = Angular frequency of magnetization, rad/sec.
 x = Distance from the centre of an electrical sheet, cm.
 A = Cross-sectional area of a test specimen, cm².
 B = Mean flux density in electrical sheet, gauss.
 B_m = Amplitude of mean flux density, gauss.
 B_s = Flux density appropriate to a state of magnetic saturation, gauss.
 H = Magnetic field strength, oersteds.
 H_e = Eddy-type component of H_0 .
 H_h = Hysteresis component of H_0 .
 H_0 = Value of H when B is zero.
 M = Intensity of magnetization.
 M_s = Saturation intensity of magnetization.
 N = Number of turns of magnetizing coil.
 P = Calculated eddy-current loss, watts.
 P_e = Actual eddy-type loss, watts.
 P_h = Hysteresis loss, watts.
 R = Resistance of a resistor, ohms.
 R_x = Resistance of an eddy-current path in the sheet, ohms.
 V_h = Voltage which is a measure of the potential drop in a resistor carrying a magnetizing current.
 V_b = Induced e.m.f. in a search coil embracing a test specimen, volts.
 V_i = E.M.F. induced in an elemental path in the sheet, volts.
 V_m = Mean amplitude of a signal comprising V_0 and a function of V_b , volts.
 V_0 = Amplitude of a signal having a square waveform, volts.
 V_{p0} = Strength of a pulse-type signal, volts.
 V_{pm} = Strength of a further pulse-type signal, volts.
 V = A voltmeter reading, volts.

(1) INTRODUCTION

In 1935 Landau and Lifshitz¹ first discussed the relationship between the velocity of a ferromagnetic domain wall and the

magnetizing field. In 1950 Williams, Shockley and Kittel² measured the velocity of a domain boundary in a single crystal of silicon iron having a known domain configuration. They showed that the motion of the boundary was retarded by eddy currents, and were able to calculate this retardation effect in terms of the domain configuration. A small discrepancy was attributed to the relaxation effects described by Landau and Lifshitz.* The exact calculation of eddy-current effects based upon a known domain configuration was compared with the classical eddy-current calculations based upon an assumed homogeneity of material, and an eddy-current anomaly factor of three could be explained in these terms. It is the present object to investigate such an explanation to determine the extent to which it is applicable to electrical sheet steels in common use.

(2) THE DOMAIN STRUCTURE OF SHEET STEELS

(2.1) The Domain Pattern

The specimens used by Williams, Shockley and Kittel had the rectangular picture-frame shape shown in Fig. 1. They were cut from a single crystal so as to have edges parallel to the

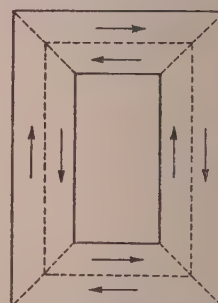


Fig. 1.—Magnetic domain structure in rectangular core cut from a single crystal of iron and having sides parallel to easy directions of magnetization.

easy directions of magnetization. There are eight domains in the specimen shown in the illustration. The broken lines indicate the positions of the domain walls and the arrows show the direction of the magnetization within the individual domains. When the domain walls occupy the positions shown the whole specimen may be said to be unmagnetized. If, however, the walls are displaced by the application of a magnetic field, the net magnetic flux in the specimen will not remain zero and the specimen will become magnetized.

Photographs of the domain patterns in a steel sheet have been obtained by Bozorth.⁴ A typical pattern is that shown in Fig. 2. It applies to a sheet of 4% silicon iron having large crystals which extend across the thickness of the sheet, this particular view being appropriate to a crystal having axes as indicated. The domain pattern shown may be regarded as representative of the

* Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Aspdén is with the English Electric Co., Ltd.

* Galt³ has since investigated the motion of a ferromagnetic domain boundary in Fe₃O₄, a material in which the eddy-current effects are small, and has provided experimental confirmation of the theory of the relaxation effects suggested by Landau and Lifshitz.

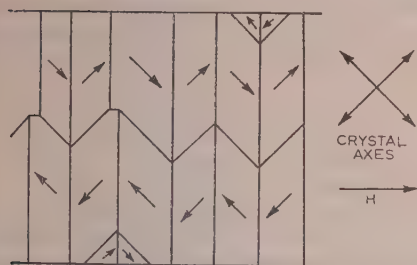


Fig. 2.—Typical domain structure in sheet-steel lamination having large crystal structure.

domain structure in sheet steels composed of large crystals. When a magnetic field H is applied to the sheet, in the direction shown, a resultant flux is established in the same direction by a movement of the 180° boundary walls which separate domains on one side of the sheet from those on the other side. The particular pattern of the domain configuration at a part of the sheet depends upon the orientation of the crystal at that point. If the crystal axes are parallel and perpendicular to the face of the sheet we may expect a very simple domain pattern of the form shown in Fig. 3 to apply. When a magnetic field is applied

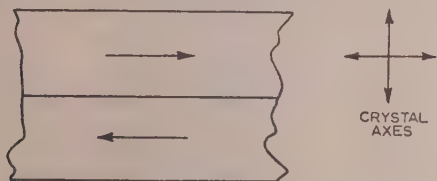


Fig. 3.—Simple domain structure in sheet-steel laminations at a part having a crystal with two of its easy axes of magnetization aligned with the surface of the sheet.

the orientation of the crystal with respect to the field direction also becomes a factor in the determination of the domain configuration.

In spite of these complexities, consideration shows that from the point of view of eddy-current effects a representative pattern of that shown in Fig. 3. On this basis the eddy-current effects in a sheet comprising two domains separated by a single plane boundary will be calculated.

(2.2) The Calculation of Eddy-Currents

It is assumed that complete homogeneity of electrical properties exists throughout the sheet. In Fig. 4 a sheet of thickness $2b$

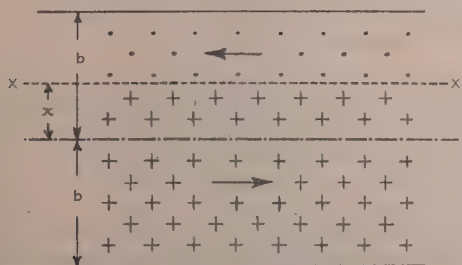


Fig. 4.—Simple domain structure having a boundary XX displaced by applied magnetic field.

shown to have a single domain boundary XX distant x from the centre of the sheet. It is assumed that a magnetic field of strength H is applied in a direction perpendicular to the plane

of the paper, directed into the paper, so that a net magnetic flux may be assumed to be established in the sheet in the same direction. A net magnetic flux exists because each domain is magnetized to saturation, the upper domain being magnetized in the opposite direction to the lower domain, and since the boundary XX is displaced from the centre of the sheet, one domain is greater than the other.

When the boundary XX is in motion such that x is increasing, currents will flow through the sheet in opposite directions on the different sides of the boundary, their directions being as indicated by the arrows. The same current must flow on one side of the boundary as flows on the other. The effective permeability within the domains is unity* since they are in a state of magnetic saturation. Thus, in a normal-size sheet magnetized at a power frequency the magnetic skin effects within the domains can be neglected. It may therefore be presumed that the current passing on each side of the boundary XX is uniformly distributed across the relevant domain. The two domains are effectively connected in series so far as induced eddy-currents are concerned, and their combined resistance per square centimetre of sheet is given by

$$R_x = \frac{\rho}{b+x} + \frac{\rho}{b-x} = \frac{2b\rho}{b^2-x^2} \text{ ohms} \quad (1)$$

Using B_s as a flux density appropriate to a state of saturation,† the electromotive force V_i induced in this path per square centimetre of sheet is:

$$V_i = 2B_s \left(\frac{\partial x}{\partial t} \right) \times 10^{-8} \text{ volts} \quad (2)$$

where t is time. The eddy-current loss per square centimetre sheet is V_i^2/R_x . If this loss is designated P , from eqns. (1) and (2):

$$P = 2B_s^2 \frac{b^2 - x^2}{b\rho} \left(\frac{\partial x}{\partial t} \right)^2 \times 10^{-16} \text{ watts} \quad (3)$$

The relationship between x and the mean flux density B in the specimen is

$$x = b \left(\frac{B}{B_s} \right) \text{ cm} \quad (4)$$

In calculating eddy-current loss it is normal to assume that B varies sinusoidally with respect to time. B may therefore be expressed in the form $B_m \sin \omega t$, where B_m is the amplitude of the mean flux density. By introducing this expression into eqn. (4) and using this equation to eliminate x from eqn. (3) the average value of P over a complete cycle of magnetization may be shown to be

$$P = \left(1 - \frac{B_m^2}{4B_s^2} \right) \frac{B_m^2 \omega^2 b^3}{\rho} \times 10^{-16} \text{ watts} \quad (5)$$

The corresponding eddy-current loss calculated in the classical manner on the basis of a uniform permeability throughout the sheet is $(B_m^2 \omega^2 b^3 / 3\rho) \times 10^{-16}$ watts. Comparing this with P , the theoretical value of eddy-current loss to be expected with an actual domain configuration, it is found that P exceeds the classical loss by a factor of $3(1 - B_m^2/4B_s^2)$.

An anomaly factor as great as three is therefore to be expected in sheet steels having domain configurations as represented in Figs. 2 and 3.

* Flux density B is given by the expression $4\pi M + H$ where M is the intensity of magnetization produced by the field H . In a saturated medium M has a fixed value M_s , say, and the value of $\partial B / \partial H$ becomes unity. It is this differential value of permeability which is effective when eddy-currents within the medium are concerned.

† Since a state of magnetization exists within a domain even in a zero field, $4\pi M_s$ may be regarded as a flux density appropriate to the state of magnetization and conveniently denoted B_s . However, an applied field which is not zero is necessary to promote boundary movement and, in deriving the simple expression for V_i in eqn. (2), it is necessary to assume that the magnetizing field is negligibly small in comparison with B_s .

A point which has been overlooked in the above treatment is that eqn. (4) applies only to the system shown in Fig. 3. To apply to the system shown in Fig. 2 it must be modified to the form

$$x = \sqrt{(2)b} \left(\frac{B}{B_s} \right) \text{ cm} \quad (6)$$

The reason for this is that a movement of the 180° domain boundaries in a direction perpendicular to that of the field corresponds to $1/\sqrt{2}$ times this movement in a direction normal to the boundary surfaces, because the 180° boundaries are inclined at 45° to the field direction. To produce a given flux change in the field direction this entails a boundary movement perpendicular to the field direction greater than is required in the system shown in Fig. 3. There must also be a corresponding modification in eqn. (2) by which the expression for V_i is reduced by the factor $1/\sqrt{2}$. The effect of these modifications is equivalent to a reduction of B_s by the factor $1/\sqrt{2}$. Consequently, the anomaly factor appropriate to the domain configuration shown in Fig. 2 becomes $3(1 - B_m^2/2B_s^2)$ instead of having the value $3(1 - B_m^2/4B_s^2)$ which applies to Fig. 3.

In this way it may be shown that the anomaly factor which we may expect to arise from domain configurations having similar characters to those shown in Figs. 2 and 3 is only slightly dependent upon crystal orientation at low or medium values of B_m . Ideally, the anomaly factor should be close to 3.0.

The uncertainties which arise from crystal orientation can be avoided, however, if attention is restricted to a consideration of the value of the anomaly factor as it applies instantaneously when, during a particular cycle of magnetization, B has the value zero. Such considerations, when put to the test in sheet steels having large crystals, afford a convenient basis for verifying the theory presented.

From eqns. (3) and (4) it may be shown that the rate of eddy-current loss when B is zero in the systems shown in Figs. 2 and 3 is given, theoretically, by

$$P = \frac{2b^3}{\rho} \left(\frac{\partial B}{\partial t} \right)^2 \times 10^{-16} \text{ watts} \quad (7)$$

This is exactly three times the classical rate of loss. The anomaly factor applicable at an instant when B is zero is therefore 3.0 and is independent of B_s , B_m , or the domain configuration.

(3) EXPERIMENTAL STUDY

(3.1) Experimental Requirements

It is required to measure the instantaneous rate of loss of the eddy-current type at the point in a cycle of magnetization at which the mean flux density is zero. The tests must be carried out on a single sheet.

The actual rate of loss has two components, one attributable to hysteresis and one attributable to eddy currents. It may be shown that the total loss rate at the instant when B is zero is

$$P_h + P_e = \frac{10H_0}{4\pi} 2b \left(\frac{\partial B}{\partial t} \right) \times 10^{-8} \text{ watts} \quad (8)$$

H_0 may be resolved into a component H_h which is independent of frequency and a component H_e which varies with frequency. The loss components may then be separated to give

$$P_e = \frac{10H_e}{4\pi} 2b \left(\frac{\partial B}{\partial t} \right) \times 10^{-8} \text{ watts} \quad (9)$$

Since this expression represents the actual eddy-type loss, it follows that the observed anomaly factor is obtained by dividing this value of P_e by the value of P as given by eqn. (7) and multi-

plying by three to base the factor upon a classical value of eddy-current loss. The experimentally observed value of the anomaly factor (a.f.) may therefore be expressed as

$$(\text{a.f.}) = \frac{30\rho H_e}{4\pi b^2} \frac{10^8}{\left(\frac{\partial B}{\partial t} \right)} \quad (10)$$

(3.2) Experimental Technique

Eqn. (10) forms the basis of an experimental measurement of the anomaly factor applicable at the appropriate instant during a cycle of magnetization. The value of $\partial B/\partial t$ is readily determined by measuring the e.m.f. induced in a search coil at the instant when the integral of the e.m.f. is zero. The corresponding value of H_e is determined by measuring the magnetizing current at the same instant, and carrying out a test somewhat similar to the method of deducing eddy-current loss from measurements of the total magnetization loss at different frequencies. The basic features of the technique and the instrument used to measure the anomaly factor on the basis of eqn. (10) are described in the Appendix.

(3.3) Experimental Results

A number of well-annealed sheet samples of steels of high silicon content were tested. Typical results are those presented in the following Table 1.

Table 1

MEASURED ANOMALY FACTORS

Steel	B_m	Anomaly factor
"48 Quality" 3½% Si $\rho = 61.3$ $2b = 0.406 \text{ mm}$	6200	3.16
	8250	3.51
	10300	3.51
Stalloy 4% Si $\rho = 59.0$ $2b = 0.300 \text{ mm}$	7000	5.91
	10500	4.26
	14000	3.85
Crystalloys 3% Si $\rho = 41.0$ $2b = 0.288 \text{ mm}$	7250	3.70
	10900	3.97
	14500	3.76
Superstalloy 4.3% Si $\rho = 59.0$ $2b = 0.309 \text{ mm}$	6775	4.81
	10250	5.22
	13550	5.03

ρ , which was measured in each individual sheet tested, is expressed in microhm per centimetre cube.

The silicon contents quoted are nominal.

The above observations provide the surprising result that the anomaly factor invariably exceeds the predicted value of 3.0. Moreover, these values of the factor are greater than the anomaly factor normally measured, i.e. the mean anomaly factor through out a complete cycle of magnetization. In the above specimen this factor is between 2.0 and 3.0. In the sample of "48 Quality" steel tested above the loss anomaly factor, as normally measured was 2.6 for B_m equal to 6200 gauss. This discrepancy is in qualitative agreement with the theory developed, since it has been shown that the anomaly factor falls with increasing flux density. However, the theory presented affords no explanation for anomaly factors in excess of 3.0. Such high anomaly factors have been recorded by observers measuring the mean effect over a complete cycle of magnetization. For example, Dannatt⁵ carried out

experiments on ring stampings representing a wide range of variables as involved in eddy-loss calculations and measured anomaly factors varying between 2.8 and 3.5.

(4) TIME-LAG EFFECTS

An anomaly factor in excess of 3.0 cannot be explained in terms of the magnetic inhomogeneity attributable to the existence of magnetic domains. The observations recorded above show that at low flux densities during a cycle of magnetization the anomaly factor is appreciably greater than 3.0, but evidence from mean loss measurements indicates that the anomaly factor falls at higher flux densities. It follows from this that some other factor is contributing a component to the eddy-type loss. There is little doubt that this factor arises from time-lag effects associated with domain boundary movements. The movement of the domain boundary walls is discontinuous and takes place in a series of jerks rather than smoothly. During their movement subsidiary eddy-current effects occur in the locality of the boundary and loss is associated with these localized eddy-currents. This loss is the hysteresis loss. The demagnetizing effect of these eddy-currents retards the motion of the boundary and causes it to move relatively slowly. Time-lags of the order of 10^{-4} sec may be associated with this movement, and these lead to an increased hysteresis loss with increasing frequency when the value of maximum flux density is maintained constant. From qualitative considerations, we can expect the time-lag to be greater if the hysteresis loss is smaller.

The samples tabulated above have a low hysteresis loss and, at a frequency of 50 c/s, the total eddy-type loss has normally a similar magnitude. We may therefore expect the time-lag to be relatively high and the observations presented above suggest that time-lag effects account for between, say, 20 and 50% of the total eddy-type loss when the flux density is zero. On this basis, it is of interest to predict the importance of time-lag in a specimen having a greater hysteresis loss and, by using a thick specimen, an eddy-type loss which is appreciably greater than the hysteresis loss. Clearly, time-lag is less and its effect upon the eddy-type loss is much more than proportionally less. This will be considered experimentally below.

To conclude this discussion of time-lag effects, it is to be noted that the time-lag is associated with boundary movements. The movements are of greater significance at low flux densities. At high flux densities the boundaries move faster, since greater field strengths and rates of hysteresis loss are involved. Thus, at high flux densities the time-lag effects become relatively unimportant and the anomaly factor falls to the normal values which can be attributed to magnetic inhomogeneities arising from domain structure.

(5) STEELS HAVING LOW ANOMALY FACTORS

(5.1) Experimental Evidence of Low Anomaly Factors

Domain configurations of the form shown in Figs. 2 and 3 are not to be found in thick sheet steels having small crystal structures. However, it seems that a sheet would have to be very thick indeed before the domains became so small relative to the thickness of the sheet that magnetic homogeneity could be taken for granted. Even then, there would be other factors which would perhaps cause the magnetic permeability to vary across the sheet. In such sheet the above discussion has shown that time-lag effects can be discounted and, with this in mind, the following further observations were obtained from a sample of Lohys steel.

Table 2

Lohys steel: 0.2% Si, $\rho = 17.1$, $2b = 0.478$ mm	
B_m	Anomaly factor
4 200	1.05
6 750	1.04
9 300	1.04

It is to be noted that the anomaly factor is substantially independent of the range of the cycle of magnetization used. This is to be expected, since the domain boundaries are likely to occupy similar positions when the flux density is zero, regardless of the cycle of magnetization. The fact that the anomaly factors were not so nearly equal with different cycles of magnetization in the previous results is attributable to a dependence of the time-lag of the boundary movement upon the cycle of magnetization.

A second feature which is apparent from the tests on the Lohys sheet is that the anomaly factor is very nearly unity. This immediately suggests that the classical eddy-current formulae apply, which indicates that the specimen is magnetically homogeneous. However, on this basis the mean loss measured over a complete magnetization cycle should also agree with classical calculation. This is not found to be the case. When Lohys sheet was tested in this respect it was found to exhibit an anomaly factor of 1.40 with B_m equal to 8 290 gauss. A possible explanation for this will now be proposed.

(5.2) Explanation of Low Anomaly Factor in Terms of Hypothetical Domain Patterns

A hypothetical domain configuration is illustrated in Fig. 5. Here, there are two domain boundaries XX and YY shown in their neutral positions to lie between the centre of the sheet and

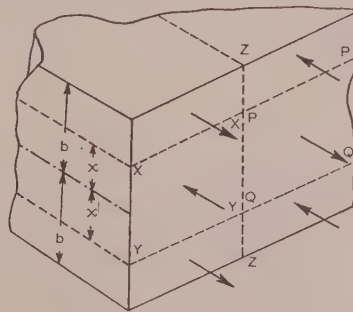


Fig. 5.—Complex domain structure in sheet steel having low anomaly factor.

a face of the sheet. Parts of the sheet between the two boundaries are assumed to be magnetized to saturation in one direction, and the remaining parts between each of the boundaries and the corresponding face of the sheet are assumed to be magnetized in the opposite direction. A boundary ZZ is shown to run transverse to the plane of the sheet. This boundary separates the three domains having boundaries XX and/or YY from three adjacent domains having corresponding boundaries PP and/or QQ, but magnetized in opposite directions, as shown by the arrows in Fig. 5. The four boundaries XX, YY, PP and QQ are distant x from the centre of the sheet in the position shown, and are regarded as having equal length. The thickness of the sheet is $2b$. Thus the resultant flux density of the system shown

is zero. Now, it may readily be shown that if the boundaries XX and YY, say, move together at the same rate, an anomaly factor of $3(b-x)/b$ is applicable. The domain configuration shown in Fig. 5 leads to an anomaly factor of unity when the flux density is zero if x has the value $2b/3$. The factors which influence domain boundary movement are such that the boundaries are moved inwards more easily than outwards initially when a magnetic field is applied. Thus, depending upon the direction of the applied magnetic field, we may expect either XX and YY or PP and QQ to move inwards before the other pair move outwards. On this basis, when the pair of boundaries moving inwards has reached a position at which x is $b/3$, the other pair of boundaries remaining with x equal to $2b/3$, the value of the anomaly factor applicable becomes 2.0. With the assumed domain configuration this corresponds to a position at which the flux density is equal to one-third of that at saturation. Thus, whereas the theoretical anomaly factor is unity when B is zero, it rises with increasing B to a sufficient extent to explain the mean loss anomaly of 1.4 as measured over a complete cycle of magnetization.

The above explanation of an example of the low anomaly effects found in some sheet steels is somewhat hypothetical, and it might appear that almost any anomaly effects within the bounds of unity anomaly factor and a factor of three can be explained by choosing a suitable domain configuration. This adds to, rather than detracts from, the significance of the explanation of the eddy-current anomaly presented.

(6) CONCLUSIONS

It is concluded that in modern electrical sheet steels anomaly factors as high as 3.0 result directly from the magnetic inhomogeneity arising from the magnetic domain configuration within the sheet. In the older-type relatively thick sheet the low anomaly factors sometimes obtained are not attributable to magnetic homogeneity as such but arise owing to particular domain configurations within the sheet. In low-loss sheet the anomaly factor is generally greater than that directly attributable to the inhomogeneities associated with the magnetic domain configuration. This is caused by a time-lag in the motion of domain boundary walls.

(7) ACKNOWLEDGMENTS

The author acknowledges a sincere debt of gratitude for the awards of a Whitworth Senior Scholarship and The Institution's I.M.E.A. Research Scholarship which enabled him to undertake this work. He is also grateful for the facilities and kind assistance afforded him at the Cambridge University Engineering Laboratories, and particularly the helpful encouragement shown him by Mr. G. E. Middleton.

(8) REFERENCES

- (1) LANDAU, L., and LIFSHITZ, E.: "Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Bodies," *Physikalische Zeitschrift der Sowjetunion*, 1935, 8, p. 153.
- (2) WILLIAMS, H. J., SHOCKLEY, W. and KITTEL, C.: "Studies of the Propagation Velocity of a Ferromagnetic Domain Boundary," *Physical Review*, 1950, 80, p. 1090.
- (3) GALT, J. K.: "Motion of a Ferromagnetic Domain Wall in Fe_3O_4 ," *ibid.*, 1952, 85, p. 664.
- (4) BOZORTH, R. M.: "Ferromagnetic Domains," *Physica*, 1949, 15, p. 207.
- (5) DANNATT, C.: "The Variation of the Magnetic Properties of Ferromagnetic Laminates with Frequency," *Journal I.E.E.*, 1936, 79, p. 663.

(9) APPENDIX—A MAGNETIZATION LOOP TESTER

The magnetization loop tester to be described was designed primarily to measure the coercive force of a magnetization loop and to study its dependence upon frequency. The loop tester is electronic, H being measured at a particular instant from a proportional voltage signal and the use of a superimposed pulse and peak measurement technique. B is measured, in effect, by gating the output from a search coil before supplying it to a rectifier-type voltmeter.

It is noted that the magnetization loop is defined as the relationship between the mean flux density B and the applied magnetic field H in a test specimen.

(9.1) Principle of Operation

The e.m.f. V_b induced in a search coil having N turns and closely embracing the test specimen is given by

$$V_b = AN \left(\frac{\partial B}{\partial t} \right) \times 10^{-8} \quad \dots \quad (11)$$

Provided that this search coil carries a negligible current, the current I in a uniformly-wound and uniformly-distributed magnetizing coil will be proportional to H and by introducing a resistor having a resistance R in series with the magnetizing coil a voltage signal V_h may be obtained which is proportional to H .

Thus
$$V_h = \frac{10}{4\pi} \frac{HR}{n} \quad \dots \quad (12)$$

The present object is to measure corresponding values of B and H or, in effect, by virtue of a suitable calibration $\int_0^t V_b dt$ and V_h .

It is presumed that the magnetization loop is not biased by a polarizing field. Thus, there are no even harmonic components in the waveforms of B or H .

Let $V_b = 0$ when $\omega t = 0$.

Consequently
$$\int_0^{\pi/\omega} V_b dt = 2ANB_m \times 10^{-8} \quad \dots \quad (13)$$

where B_m is the amplitude of B .

At the particular instant t_1 the value of B is given by

$$\int_0^{t_1} V_b dt = AN(B + B_m) \times 10^{-8} \quad \dots \quad (14)$$

Also
$$\int_{t_1}^{\pi/\omega} V_b dt = AN(B_m - B) \times 10^{-8} \quad \dots \quad (15)$$

From eqns. (14) and (15)

$$\int_0^{t_1} V_b dt - \int_{t_1}^{\pi/\omega} V_b dt = 2ANB \times 10^{-8} \quad \dots \quad (16)$$

In Fig. 6(a) typical waveforms of V_b and V_h are shown. In order to measure the difference between the two integrals in eqn. (16) a signal having a square-type waveform as shown in Fig. 6(b) is generated. The amplitude V_0 of this waveform is greater than that of the signal V_b . A leading edge occurs at the instant t_1 . The square wave is added to the signal V_b , as shown in Fig. 6(c), and the composite signal is rectified, as shown in

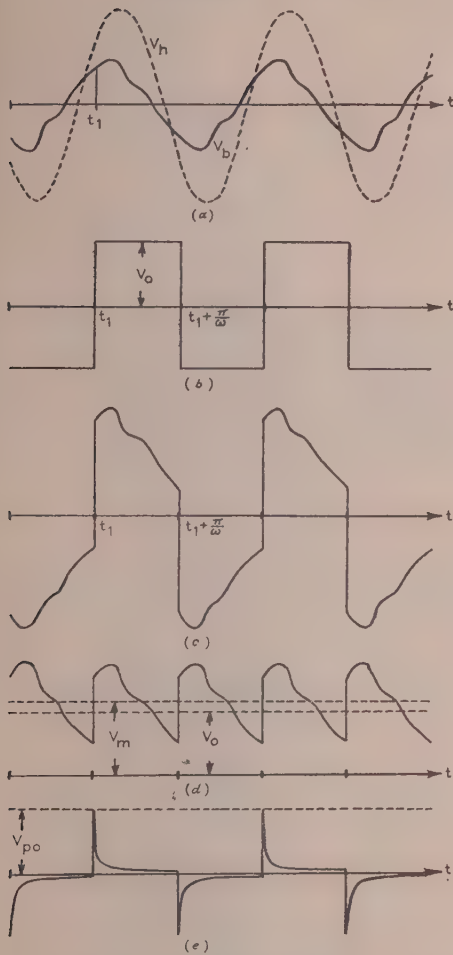


Fig. 6.—Signal waveforms illustrating operation of magnetization loop tester.

(a) Typical waveforms of V_b and V_h .

(b) Square-type waveform generated in order to measure the difference between the two integrals in eqn. (16).

(c) Square wave added to signal V_b .

(d) Waveform (c) rectified.

(e) Pulse-type signal generated, in order to measure value of V_h at instant t_1 , by signal having square-type waveform (b).

Fig. 6(d). The mean value V_m of this rectified signal is equal to V_0 plus

$$\frac{\omega}{\pi} \left(\int_{t_1}^{t_1 + \frac{\pi}{\omega}} V_b dt - \int_0^{t_1} V_b dt \right)$$

Thus

$$\int_0^{t_1} V_b dt - \int_{t_1}^{t_1 + \frac{\pi}{\omega}} V_b dt = \frac{\pi}{\omega} (V_0 - V_m) \quad (17)$$

Accordingly, from eqns. (16) and (17)

$$B = \frac{10^8 \pi}{2AN\omega} (V_0 - V_m) \quad (18)$$

In order to measure the value of V_h at the instant t_1 the signal having the square-type waveform shown in Fig. 6(b) is used to generate the pulse-type signal shown in Fig. 6(e), the amplitude of the pulses being V_{p0} . When this signal is added to the signal

V_h , a measurement of the peak voltage V_{pm} will be equal to the sum of V_{p0} and V_h at the instant t_1 . Thus, from eqn. (12)

$$H = \frac{4\pi n}{10R} (V_{pm} - V_{p0}) \quad (19)$$

(9.2) The Instrument

In Fig. 7 the circuit of a magnetization loop tester operating on the above principles is shown. The test specimen C is magnetized by current flowing in the coil P. This coil is connected in series with a resistor R and the whole is supplied from the variable-frequency source E. There are two identical secondary windings formed by the search coils S_1 and S_2 . Each of these has N turns and affords an output signal of strength V_b .

The source E also supplies a phase-changer F. This phase-changer has the form of an inductor connected in series with a variable resistor, the output being drawn from across the resistor. However, if it is not simply the coercive force of the magnetization loop that is to be measured, a more complex phase-changer affording a very wide phase adjustment may be required.

The output from the phase-changer is fed to a squaring amplifier formed by the valve V_1 . The potentiometer R_1 enables the "mark space" ratio of the square wave to be adjusted to unity. The amplitude of the square wave is determined by the limiting device formed by the valve V_2 . The potentiometer R_2 serves to adjust this amplitude.

The resistor-capacitor circuit RC_1 serves to bring the d.c. component of the signal having the square-type waveform to a fixed level. The RC value of the circuit RC_1 is very high so that it may pass the a.c. component of the square-wave without distortion. However, a similar resistor-capacitor circuit RC_2 serves the dual purpose of bringing the d.c. component to the fixed level and distorting the square wave into a pulse-type waveform as shown in Fig. 6(e).

The signal V_b supplied by the search coil S_1 is added to the output from the circuit RC_1 and the composite signal is supplied to the grid of the valve V_3 , which acts as a cathode-follower. The cathode potential is therefore proportional to the sum of the above-mentioned fixed d.c. level, the square-wave signal, and the signal V_b . By using a rectifier-type voltmeter, M_1 , to compare the cathode potential of the valve V_3 with the potential of the potentiometer tapping T_1 , and by suitably designing this potentiometer, the effects of the fixed d.c. level can be compensated and the voltmeter M_1 can be arranged to respond solely to a signal having a square-wave component and the V_b component. When calibrated, this voltmeter M_1 gives a direct measure of the factor $(V_0 - V_m)$ of eqn. (18).

The signal V_h which is supplied by the resistor R is added to the pulse-type output of the circuit RC_2 and then fed to the peak voltmeter formed by the valve V_4 , the capacitor C_1 , the cathode-follower valve V_5 and the voltmeter M_2 . A potentiometer tapping T_2 provides a reference potential which compensates for the fixed d.c. level of the signal supplied by the circuit RC_2 . The voltmeter M_2 may be calibrated to give a direct measure of the factor $(V_{pm} - V_{p0})$ of eqn. (19).

If desired, the two potentiometer tappings T_1 and T_2 may be adjusted to provide potentials which compensate for V_0 and V_{p0} respectively. This facilitates the calibration of the voltmeters.

Calibration is effected by replacing the signals V_b and V_h by known d.c. signals.

The value of B_m can be adjusted by varying the voltage of the source E. It may be measured by utilizing an additional cathode-follower circuit including the valve V_6 . The rectifier-type voltmeter M_3 can be calibrated to give a direct measure of V , where

$$B_m = \frac{10^8 \pi}{2AN\omega} (V) \quad (20)$$

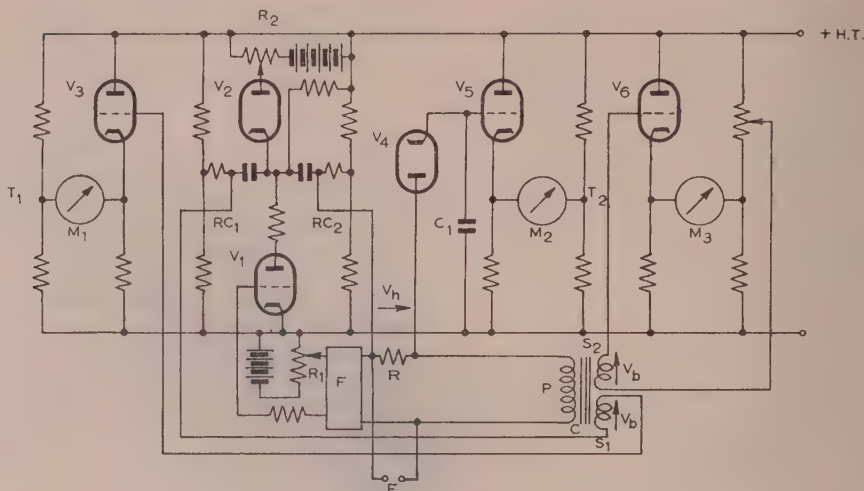


Fig. 7.—Circuit of magnetization loop tester.

(9.3) Application of the Instrument

When applied for the purpose of measuring the anomaly factor as given in eqn. (10), the magnetization loop tester must also be used to measure the rate of change of B at the point on the magnetization loop at which B is zero. This is readily achieved by incorporating in the circuit shown in Fig. 7 a suitable switching system by which the signal V_h supplied to the electronic circuit can be replaced by the signal V_b supplied by the search coil S_2 . To avoid complication of the circuit this switching system is not shown in the Figure.

(9.4) Operation and Results

Once the instrument has been adjusted and calibrated the testing of a sheet specimen is quite straightforward. It is desirable to use specimens having a standard magnetic length. Thus rectangular specimens 22 cm long by 7 cm wide were used, and a central slot 1.27 cm wide and 18 cm long was cut in each of them. This gave a mean magnetic length of 53.3 cm when the specimen was magnetized around the rectangular path formed, and in conjunction with the calibration of the instrument this was found to give the measurement of field strength by the voltmeter M_2 as 1 oersted/volt. From a knowledge of the dimensions of the test specimen there is little difficulty in applying eqn. (20) and calibrating the voltmeter M_3 so that B_m can be measured. When a particular value of B_m has been chosen at which to conduct a test, it is simply a matter of adjusting the source E to give the required frequency and a suitable output voltage. To measure the coercive force of the magnetization loop the phase-changer is then adjusted until the voltmeter M_1 gives a null indication. From eqn. (18) B is then zero. M_2 is read, both when a signal V_h is applied to it and when the circuit is adjusted so that a signal V_b is applied to it. The calibration constant in the latter case (previously determined as 4.3 volts/volt) enables the value of V_b at the appropriate instant to be measured. Tests carried out at different frequencies but using the same value of B_m enable the coercive force to be measured as a function of frequency. In general, a linear relationship between coercive

force and frequency was obtained, and the frequency-dependent component could readily be deduced. A typical example of the readings afforded by the voltmeter M_2 is given below. The results apply to a Superstalloxy specimen tested with B_m equal to 13 550 gauss. From these results the eddy-type component of field strength has an estimated value of about 0.33 oersted at 50 c/s. The corresponding value of V_b is 3.15 times the calibration factor 4.3, or 13.5 volts.

Table 3

Frequency	V_h	V_b
c/s	volts	volts
10	0.55	0.41
20	1.29	0.48
30	2.00	0.54
40	2.54	0.61
50	3.15	0.68

$$\text{From eqn. (10)} \quad \text{a.f.} = \frac{30\rho H_e}{4\pi b^2} \frac{10^8}{\left(\frac{\partial B}{\partial t}\right)} \quad (21)$$

Hence, from eqns. (11) and (21)

$$\text{a.f.} = \frac{30AN\rho H_e}{4\pi b^2 V_b} \quad (22)$$

From the data of the specimen and the above results

$$\begin{aligned} 2b &= 0.0309 \text{ cm} \\ A &= 0.0883 \text{ cm}^2 \\ N &= 4000 \\ \rho &= 59.0 \text{ microhm-cm} \\ H_e &= 0.33 \text{ oersted} \\ V_b &= 13.5 \text{ volts} \end{aligned}$$

When these data are substituted in eqn. (22) the anomaly factor found to be applicable is 5.03.

MAGNETIC TIME-LAG EFFECTS IN SOLID STEEL CORES

By H. ASPDEN, Ph.D., B.Sc., Wh.Sc., Graduate.

(The paper was first received 14th September, and in revised form 7th November, 1955. It was published as an INSTITUTION MONOGRAPH in January, 1956.)

SUMMARY

The discrepancy between the theoretical eddy-current effects and the actual eddy-current effects in thick steel cores is investigated. By using thick cores the magnetic inhomogeneity arising from ferromagnetic domain configuration is of little consequence and the discrepancy may be attributed to an intrinsic time-lag effect in the magnetization process. This time-lag is measured and explained in terms of a localized eddy-current action associated with the dissipation of hysteresis loss. It was found that the time-lag depended upon frequency and was less important at high frequencies.

LIST OF PRINCIPAL SYMBOLS

- τ = A time constant associated with a magnetization time-lag, sec.
 μ = Permeability.
 μ_0 = The modulus of a complex permeability.
 θ = The argument of a complex permeability, rad.
 H = Magnetizing field strength, oersted.
 H_n = The amplitude of a varying field strength, oersted.
 ω = Angular frequency of magnetization, rad/sec.
 B_m = Amplitude of a varying flux density, gauss.
 μ_h = Modulus of a complex permeability representing static hysteresis loop.
 θ_h = Argument of a complex permeability representing static hysteresis loop, rad.
 z = A factor defined in eqn. (7).
 ρ = Resistivity, ohm-cm.
 H_D = A polarizing magnetic field strength, oersteds.
 a = The radius of a test specimen, cm.
 Q = The ratio of a mean flux density at a particular frequency to the amplitude of the flux density at a surface at zero frequency.
 D = A distance between adjacent ferromagnetic domain walls, cm.
 d = The distance through which a domain wall moves, cm.
 B_s = Saturation flux density, gauss.
 \bar{B} = Flux density, gauss.
 t = Time, sec.
 x = Displacement of a boundary wall, cm.
 H_{coer} = A coercive force, oersteds.

(1) INTRODUCTION

It is well known that there is a discrepancy between the observed and calculated values of eddy-current loss in electrical sheet steels at supply frequencies. It is found convenient to denote the ratio between an observed eddy-type effect and a calculated one by the term "anomaly factor."

Brailsford¹ has considered the extra loss which arises owing to the flux waveform distortion produced by the non-linear character of the magnetization loop within individual sheets and from sheet to sheet of a laminated core, and the increased hysteresis loss

due to eddy-current skin effects as demonstrated by a model. In this way he was able to explain a substantial portion of the extra loss observed in specimens of low-silicon and high-silicon steels which gave loss anomaly factors of about 1.5 and 2.0 respectively. He also describes an experiment on a low-silicon steel disc in a rotating field the results of which are interpreted as indicating the absence of an anomaly.

At higher frequencies of magnetization no direct agreement between observed losses and those calculated from simple formulae can be expected, and eddy-current screening must be considered. An accurate mathematical approach to this problem is extremely difficult, but certain simple assumptions lead to solutions which indicate roughly the nature of the result. A useful approach is to employ the model lamination as devised by Brailsford,¹ or a more simple type developed by Heys.² At high frequencies it is usual to consider the variation in the effective permeability rather than the loss, and it is known that an anomaly is manifested under these conditions by a more rapid fall in permeability as frequency is increased than theory would predict. However, when model tests are carried out and the high-frequency conditions in one lamination are simulated by a group of laminations at low frequency it is found that in the actual case the permeability falls rather more rapidly than would be expected from tests on a model.

Stewart³ has attributed the eddy-current anomaly to a time-lag between the magnetizing field and the flux density, a time-lag, that is, which has some intrinsic character and is independent of classical eddy-current action. This conception affords a ready explanation of the eddy-current anomaly. Landau and Lifshitz⁴ have shown in a classic paper on domains that time-lag effects apart from eddy-current action have some physical basis in ferromagnetism.

The above explanations of the eddy-current anomaly virtually neglect the non-homogeneity of the electrical and magnetic properties of steel. This can reasonably be justified by a careful preparation and treatment of a test specimen, and provided that the dimensions of the specimen are large in comparison with the size of the ferromagnetic domains within the specimen. Ferromagnetic domains may have a width of 0.1 mm, and since modern low-loss electrical sheet steel has a comparable thickness magnetic homogeneity cannot be the basis of an accurate eddy-current calculation.

However, it has been shown by the author⁵ that the eddy-current anomaly in low-loss sheet cannot be wholly attributed to magnetic inhomogeneity arising from the ferromagnetic domain configuration within the sheet. The suggestions of Brailsford, Stewart, and Landau and Lifshitz therefore merit reconsideration, since it seems feasible to suppose that inhomogeneity arising from domain structure contributes part of the anomaly effect and some other causes, such as waveform distortion and time-lag effects, contribute the remaining part of the anomaly effect.

This paper deals with an investigation of the latter effects by carrying out tests on solid steel specimens which are sufficiently thick for the problem of domain configuration to be ignored and the specimen to be regarded as magnetically homogeneous.

Correspondence on Monographs is invited for consideration with a view to publication.
 Dr. Aspden is with the English Electric Co., Ltd.

The existence of a fundamental time-lag effect in magnetization is assumed. It is also assumed that this effect can reasonably be represented by a time-constant which will be denoted τ and which will be assumed to be constant independent of the flux density in the specimen.

(2) THE DYNAMIC HYSTERESIS LOOP

Tests on electrical sheet steel specimens have shown that under alternating magnetization at 50 c/s the hysteresis loop is modified from its static form. When a magnetizing field having a fixed amplitude is used for these tests the modification has a form such as that shown in Fig. 1. This modification can readily be explained in terms of a time-lag effect.

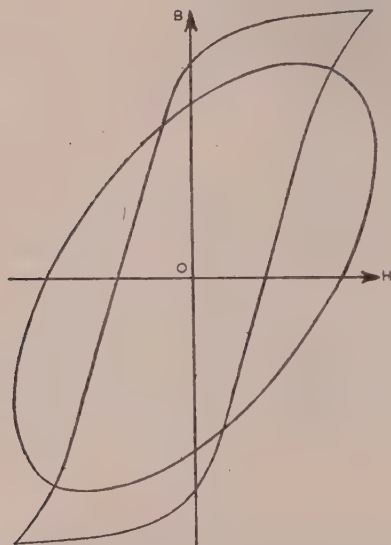


Fig. 1.—Static hysteresis loop and corresponding dynamic hysteresis loop.

Whereas the static hysteresis loop has a well-known shape, the dynamic hysteresis loop is somewhat elliptical. Thus, if an electrical specimen is tested and, owing to the existence of a time-lag effect, the hysteresis loop is everywhere modified to an elliptical form, one would not expect there to be any non-linearities of a form which would cause a waveform distortion. Purely sinusoidal changes of the applied magnetizing field-strength may reasonably be expected to be accompanied by corresponding sinusoidal changes of the resultant flux passing through the specimen. On this basis, any anomaly effect observed cannot be attributed to the waveform distortion effects discussed by Brailsford, and substantially the whole effect measured can be attributed to a time-lag. It must be emphasized that this argument presupposes the existence of a time-lag effect.

(3) THE EFFECT OF TIME-LAG UPON EDDY-CURRENT THEORY

In studying the effects of a time-lag upon eddy-current theory it is convenient to assume that the hysteresis loop has adopted an elliptical form, and to utilize the conception of a complex permeability of the form

$$\mu = \mu_0 e^{-j\theta} \quad (1)$$

It follows that if the magnetizing field H has the form $H_n \sin \omega t$, the flux density B will have the form

$$B = \mu_0 H_n \sin(\omega t - \theta) = B_m \sin(\omega t - \theta) \quad (2)$$

where B_m is the amplitude of B . By eliminating the parameter ωt from the expressions for B and H , the B/H relationship may be shown to be an ellipse of the form shown in Fig. 2. The value of the semi-major axis of the ellipse is approximately

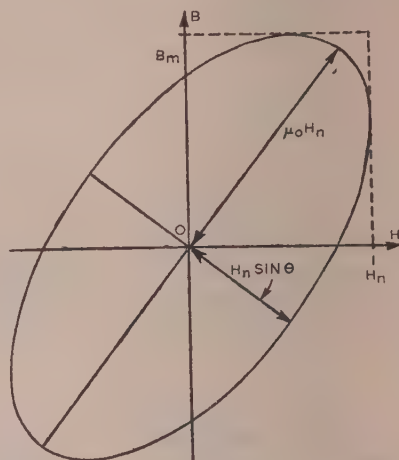


Fig. 2.—Hysteresis loop based upon a pure complex permeability $\mu_0 e^{-j\theta}$.

$\mu_0 H_n$ and the value of the semi-minor axis is approximately $H_n \sin \theta$. The area of the loop formed by the ellipse is therefore $\pi \mu_0 H_n^2 \sin \theta$.

The conception of the complex permeability as a means for simulating hysteresis has been used by Latour⁶ in the calculation of magnetization losses in sheet steels.

Now, if hysteresis is ignored for a moment and the effects of a time-lag are considered, the following equation may be used to relate B and H :

$$\mu_h H - B = \tau \frac{\partial B}{\partial t} \quad (3)$$

If H is put equal to $H_n \sin \omega t$ as before, and if it is assumed that the permeability μ_h is constant, the effective permeability μ at the angular frequency ω may be shown from eqn. (3) to be given by

$$\mu = \frac{B}{H} = \frac{\mu_h}{\sqrt{1 + \omega^2 \tau^2}} e^{-j \arctan \omega \tau} \quad (4)$$

This equation is analogous with eqn. (1), and it therefore appears that the conception of a complex permeability can also be used to account for time-lag effects.

To use a complex permeability to simulate both hysteresis and time-lag effects the values of μ_0 and θ in eqn. (1) should be

$$\mu_0 = \frac{\mu_h}{\sqrt{1 + \omega^2 \tau^2}} \quad (5)$$

$$\theta = \arctan(\tan \theta_h + \omega \tau) \quad (6)$$

where μ_h and θ_h have values such that $\mu_h e^{-j\theta_h}$ represents the static hysteresis loop.

From the above analysis two experimental approaches present themselves. The variation of effective permeability with frequency can be measured and eqn. (5) used to deduce the value of τ . Alternatively, we can measure the dependence of θ upon frequency and deduce τ by using eqn. (6).

(4) MEASUREMENT OF EDDY-CURRENT EFFECTS

(4.1) The Limiting Value of Phase Angle

Bethenod⁷ has shown that, for a magnetic system comprising a closed magnetic circuit of sheet-iron laminations, the limiting phase angle under high-frequency conditions between the e.m.f. applied to the magnetizing winding (assumed equal and opposite to the induced back e.m.f.) and the magnetizing current is not $\pi/4$ radians as indicated by the formulae of J. J. Thomson,⁸ but is $\pi/4$ diminished by an angle $\theta/2$ radians. If θ is the angle associated with hysteresis, i.e. the loss angle of the hysteresis loop, the limiting phase angle involved at frequencies corresponding to an appreciable degree of magnetic skin effect is $\frac{\pi}{4} - \frac{\theta}{2}$.

The author⁹ has shown that the same phase angle applies in a specimen having a solid core. It would therefore appear that the most direct way of measuring the time-lag is to measure the variation of phase angle with frequency.

It is first essential to consider the errors which may arise if there are inadequate magnetic-skin effects for the above value of

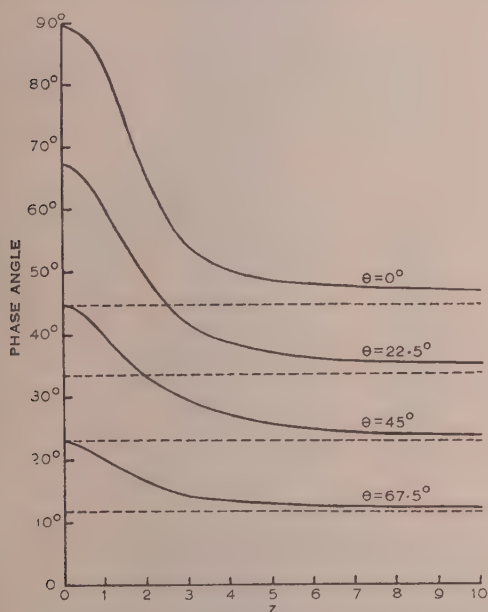


Fig. 3.—Theoretical values of limiting phase angle of magnetizing circuit due to eddy currents in magnetized core.

phase angle to be applicable. In Fig. 3 the theoretical phase angle associated with the magnetizing circuit used to magnetize a specimen of circular cross-section is shown. The phase angle is given as a function of z and θ where

$$z^2 = \frac{1}{4} \pi \omega \mu_0 a^2 \times 10^{-9} / \rho \quad (7)$$

and μ_0 and θ have values such that $\mu_0 \epsilon^{-j\theta}$ represents the dynamic hysteresis loop.

The broken lines in Fig. 3 indicate the asymptotes to which the curves tend. The error angle at the higher values of z is $\cos\left(\frac{\pi}{4} + \frac{\theta}{2}\right)$ rad. The greater z the less the error, but also,

since the magnetic-skin thickness is a/z , the greater z the less the skin thickness. It is not desirable to introduce too much skin effect, since the skin thickness may become too low in comparison with the domain size and the assumption of magnetic homogeneity

will not be justified. Hence there must be a compromise in choosing a frequency range over which measurements should be made.

The significance of the above considerations is best appreciated from the following numerical example. If an ordinary steel rod or diameter 1.3 cm having, say, a permeability of 1000 and a resistivity of 20 microhm-cm is magnetized at 200 c/s, the value of z is 18.3. This gives a magnetic skin thickness of 0.35 mm, which is appreciably greater than the dimensions of the magnetic domains present in the rod. The error angle between the actual phase angle and that represented by the asymptote of the form shown in Fig. 3 is 0.87° when θ has the value 22.5° . At 400 c/s the value of z is 25.9 and the error angle is 0.61° . It is therefore reasonable to expect a fall in the measured phase angle of about 0.26° between these frequencies, provided that θ itself is independent of frequency. Evidence of a fall greater than this would be evidence of time-lag effect of the kind mentioned above.

If, for example, τ is 10^{-4} sec and θ_h is 22.5° , eqn. (6) indicates a variation of θ from 28.4° at 200 c/s to 33.7° at 400 c/s. This corresponds to a fall of phase angle of 2.65° which is more than ten times that to be expected owing to skin-effect variations.

These considerations are important, since it is desired to detect the time-lag and measure the time-constant without knowing the appropriate values of μ_h and θ_h to use in simulating the hysteresis loop. The latter factors depend upon too many variables and cannot be predicted with certainty, but the asymptotic expression for the phase angle is independent of these factors and affords a neat experimental approach. There seems to be only one weakness involved, and this lies in the assumption that θ_h is constant. It is, in fact, dependent upon the value of B_m . However, always provided that the specimen is magnetically homogeneous, when sufficient magnetic-skin effects occur the proportions of material magnetized to particular values of B_m are the same independent of frequency, although the amount of material magnetized to a particular value of B_m varies with frequency. This tends to justify the assumption that θ_h is constant.

A further advantage of the limiting-phase-angle technique for measuring τ is that there is no restriction as to the shape of the test specimen.

A cast-steel ring was used in the measurements recorded here. The ring had a square cross-section of area 1.52 cm^2 , and the mean length of the magnetic path was 37 cm. The properties of the ring were such that the considerations applied to the rod specimen example above were roughly applicable. A test frequency range of 200–400 c/s was therefore suitable, but the tests were carried out over the range 30–500 c/s.

The measurements of the phase angle between an e.m.f. induced in a search coil and a magnetizing current were carried out using a bridge network with an oscillograph as a detector. It was not practicable to use a galvanometer for the balance indicator, owing to the harmonics present in the system, whereas a tuned galvanometer would have rendered measurements at different frequencies tedious. An oscillograph with high amplification was used to detect the balance and was found to be exceptionally suitable, since it gave a visible picture of the harmonics present and enabled an elimination of the fundamental component with reasonable assurance. In many of the measurements the 3rd harmonics were very pronounced but very much smaller than the fundamental, particularly at low magnetizing currents. This enabled a ready adjustment to be made by obtaining what appeared to be the nearest approach to a 3rd harmonic. The sensitivity obtainable under these conditions was such that the angle could be measured to within 0.1° .

Typical results obtained from tests on the ring specimen described are presented in Fig. 4. Two curves relating phase

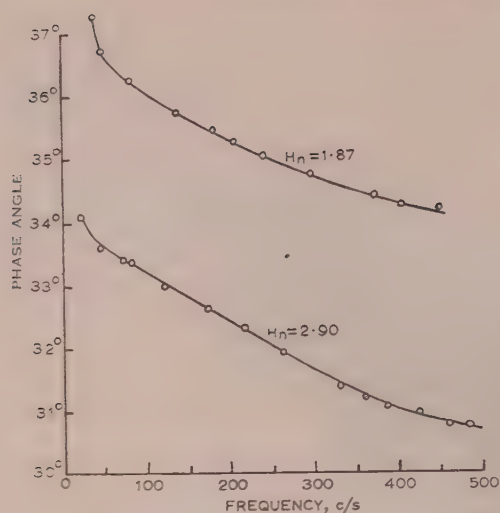


Fig. 4.—Experimentally observed phase angle.

angle and frequency are shown, each applying to a different surface magnetizing field. The steady fall of phase angle as the frequency is increased is evidence of a time-lag effect. The value of τ may be obtained by using the following equation:

$$\tau = \sec^2 \theta \frac{\partial \theta}{\partial \omega} \quad (8)$$

This is a differential equation derived from eqn. (6). Corresponding values of θ and $\partial \theta / \partial \omega$ can be obtained at a particular frequency from the curves shown in Fig. 4 and the value of τ deduced from eqn. (8). Thus, for example, at 300 c/s the results indicate that τ is 3.29×10^{-5} sec when the amplitude of the surface field strength is 1.87 oersted and 4.73×10^{-5} sec when this amplitude is 2.90 oersted. However, from the discussion of errors given above, values of τ should, in fact, be less. Probable values are approximately 2.5×10^{-5} and 4×10^{-5} sec respectively.

The curves of Fig. 4 are interesting in that they exhibit a somewhat linear fall of the phase angle over the range 100–300 c/s, but it seems that at higher frequencies the rate of fall gradually decreases. This suggests that the time-lag is decreasing as the frequency increases. This is even more apparent in the tests which will be described below.

(4.2) The Effective Permeability

It is not new to measure the effective permeability of an iron core over a range of frequencies with a view to applying eqn. (5) to determine τ . In these experiments an applied magnetizing field having an amplitude independent of the frequency is used and the e.m.f. induced in a search coil embracing the core is measured. This e.m.f., when divided by the frequency, falls with increasing frequency in exactly the same way as the effective permeability.

The mean flux density within the core varies somewhat inversely with frequency. At least a part of this variation is directly attributable to the fact that the magnetization curve of the core is not exactly linear. At low flux densities the permeability is less, and since—owing to magnetic skin effects—the flux density within the core decreases, the effective permeability falls in consequence. This effect cannot be ignored in assessing τ .

In carrying out these tests it was decided to measure the effects of a polarizing field upon the value of τ by displacing the magnetization loop. A 3-limb core construction was adopted for this purpose. Since the shape of the a.c. flux path is important in these tests, this core was formed from two long solid steel rods

of circular section arranged parallel and closely adjacent to one another with their ends connected together by yoke members. The third limb was formed by a steel bar which bridged the two yoke members. Each rod carried a magnetizing coil, and the coils were connected in series and in series with a ballast resistor, and the whole circuit was energized from a variable-frequency source producing a good sinusoidal waveform. The resistor was adjusted at different frequencies to give a particular value of magnetizing field strength. Valve voltmeters were used to measure this field strength from the p.d. across a resistor carrying the magnetizing current and the mean flux density from the e.m.f. induced in a search-coil system embracing the rods. The third limb carried a magnetizing winding which was energized from a d.c. source and used to apply the polarizing field to the rods. This field was measured in terms of the current in the magnetizing winding, the ratio between the field and the current being estimated from a magnetic circuit calculation.

Typical results obtained from tests on mild-steel rods of 5/16 in diameter are presented in Table 1.

Table 1

MEAN AMPLITUDE OF FLUX DENSITY WITH APPLIED A.C. MAGNETIZING FIELD OF AMPLITUDE 3.5 OERSTEDS AS FUNCTION OF FREQUENCY AND POLARIZING FIELD H_D

H_D	Flux density				
	At 0 c/s	At 113 c/s	At 455 c/s	At 1010 c/s	At 20 kc/s
oersteds	gauss	gauss	gauss	gauss	gauss
0	2 850	960	580	430	82
2	2 730	960	580	430	85
4	2 620	960	580	430	91
6	2 500	960	580	430	95
8	2 460	890	565	412	92
10	2 310	840	515	378	83
12	2 160	780	460	336	74
14	1 920	730	415	307	69
16	1 620	650	360	280	62
18	1 500	600	330	256	58
20	1 310	570	315	220	52

The available equipment used in the tests enabled frequencies up to about 1 kc/s and from 14 to 20 kc/s to be used.

Before the above data are analysed it is of interest to consider the theoretical effects of the time-lag upon the effective permeability.

Under conditions of appreciable skin effect the penetration function, as defined by the ratio of the amplitude of the mean flux density to the amplitude of the flux density at the surface, is $2/z$ in a solid cylindrical specimen, where z has the form given in eqn. (7).

Now, since the field strength at the surface of the specimen has a constant amplitude, eqn. (5) shows that the amplitude of the flux density at the surface will fall with frequency. The result is that the amplitude of the mean flux density within the specimen

will fall to $\frac{2}{z\sqrt{(1 + \omega^2\tau^2)}}$ times the value it has at zero frequency.

Defining Q as this factor, we have, from eqn. (7)

$$Q\sqrt{(\omega)} = \sqrt{\left[\frac{\rho \times 10^9}{\pi \mu_h a^2 \sqrt{(1 + \omega^2\tau^2)}} \right]} \quad (9)$$

This relationship is shown with τ as a parameter in Fig. 5.

The data given in Table 1 can be presented in a similar form, as shown in Fig. 6. It will be observed by comparing these curves that evidence of the value of τ is available directly. The

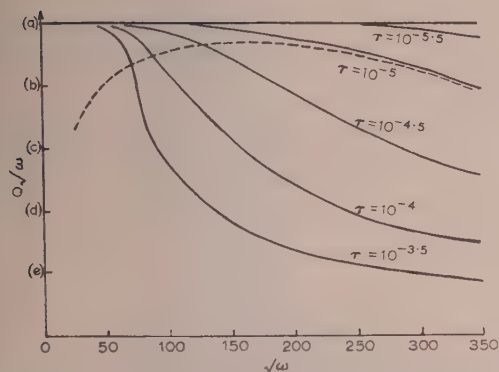


Fig. 5.—Theoretical values of Q/ω given as functions of $\sqrt{\omega}$, with τ as parameter.

Scale values:

- (a) : $\sqrt{\left(\frac{\rho \times 10^9}{\pi \mu_h a^2}\right)}$
 (b) : $0.8(a)$
 (c) : $0.6(a)$
 (d) : $0.4(a)$
 (e) : $0.2(a)$

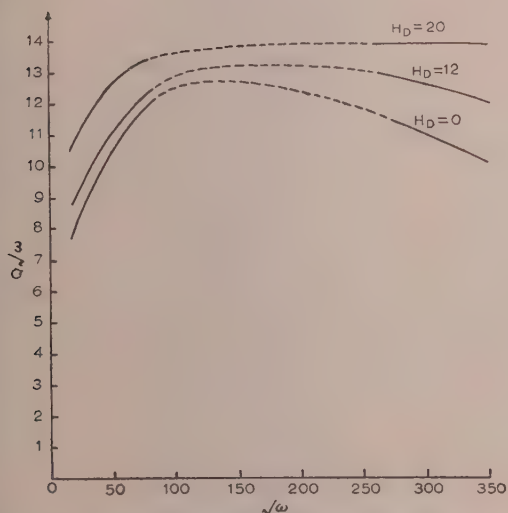


Fig. 6.—Observed values of Q/ω given as function of ω with polarizing field H_D as parameter.

Results are presented in such a way that they are independent of the actual value of permeability, provided that sufficient magnetic skin effect occurs. This is assured, since the range of frequencies covered is comparable with that used in the limiting-phase-angle experiments, although a range of much higher frequencies has also been covered. In the latter respect there has been a slight departure from the original intention. Some of the tests have been extended to a range over which the magnetic-skin thickness is theoretically about 50 microns, a distance which is comparable with the dimensions of a large magnetic domain. However, domain size in an ordinary mild-steel rod may still be sufficiently small for representative results to be obtained, and this view is substantiated by the data shown in Fig. 6.

These data indicate that τ falls with increasing frequency. For example, compare the curve $H_D = 0$ in Fig. 6 with the broken curve in Fig. 5, which applies where τ is 10^{-4} sec at 1 kc/s, 3×10^{-5} sec at 1600 c/s and 10^{-5} sec at 10 kc/s. The variation of τ with frequency can best be estimated on this basis.

The curves in Fig. 6 also indicate that the effect of a polarizing field is to decrease the value of τ , suggesting that τ varies according to the flux density in the specimen. (It has been presumed to be constant for the purpose of analysis.)

Finally, it is of interest to compare the numerical values of the effective permeability appropriate to the broken curve shown in Fig. 5 and the curve $H_D = 0$ shown in Fig. 6. A comparison of the ordinate scales indicates that $\sqrt{\left(\frac{\rho \times 10^9}{\pi \mu_h a^2}\right)}$ has the value 13.05. When the measured values of a (0.397 cm) and ρ (17.1 microhm-cm) are introduced μ_h is found to be 2040. This is appreciably greater than the value of 815 obtained by dividing 2850 gauss from Table 1 by 3.5 oersted. The reason for this is that the value of μ_h as obtained by measuring the complex permeability of the actual hysteresis loop under static conditions is 2400, and this is effectively reduced when magnetic skin effects occur, since those parts of the magnetized core near the centre are subjected to a low magnetizing field and have a correspondingly lower permeability. This will be most readily understood by reference to Fig. 7, where an analysis of the complex

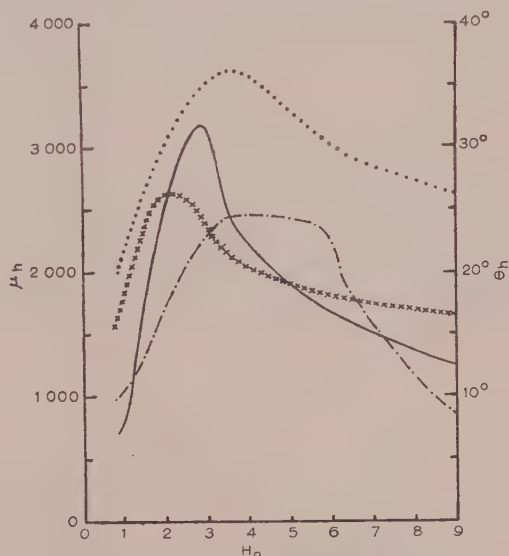


Fig. 7.—Parameters of complex permeability representing static hysteresis loops of the core tested.

- μ_h with sinusoidal H .
 - - - μ_h with sinusoidal B .
 xxxxxxxxxxxx θ_h with sinusoidal H .
 θ_h with sinusoidal B .

permeability of the test specimen is presented. The values depend upon the waveform of the magnetizing field used, and may be obtained by first plotting the hysteresis loop, then drawing the corresponding waveforms of B and H and deducing the fundamental component of the one which does not vary sinusoidally. The ratio of the amplitudes and the phase angle then afford a measure of the quantities μ_h and θ_h respectively.

Referring to Fig. 7 it will be seen that when the magnetizing field is about 3.5 oersteds the value of μ_h is the same regardless of the waveform used. It is presumed that intermediate waveforms between those corresponding to a purely sinusoidal H and a purely sinusoidal B will produce a permeability relationship intermediate to those shown in the Figure. Thus, at zero frequency the initial permeability is 2400. When appreciable magnetic skin effects occur, the field strength throughout the specimen will vary from a very low value at the centre of the

specimen to the value 3.5 oersteds at the surface. The appropriate mean value of μ_h is difficult to predict but it has been found to be 2040. This appears consistent with the curve shown in Fig. 7 since, in view of the peak of μ_h with sinusoidal H , which is experienced at a lower field strength than 3.5, the fall of μ_h to be expected from the sharp decrease of the permeability curves at low field strengths will be somewhat compensated.

The comparison effected between Figs. 5 and 6 applies only when the limiting conditions involving sufficient skin effect have been reached. Over an initial frequency range, not only do the asymptotic expressions upon which Fig. 5 is based not apply, but the value of μ_h is varying, in effect. In view of this, little weight can be placed upon the very long time-lag which appears to exist at frequencies below $\sqrt{\omega} = 50$. However, there is little doubt that sufficient skin effect exists above this frequency for the analysis to be sufficiently applicable.

(5) THE THEORY OF THE TIME-LAG

The time-lag cannot be sufficiently accounted for in terms of the relaxation mechanism suggested by Landau and Lifshitz.⁴ However, a time-lag can be associated with the motion of the domain boundary walls, this time-lag arising owing to the retardation effects of the local eddy-currents which serve to dissipate the hysteresis loss.

When a domain wall is displaced from a stable position by the application of a magnetizing field, it moves in a reversible manner until a position of instability is reached. Then, since the applied field is more than sufficient to move the wall, the wall moves suddenly to a new position of stability. The speed of this movement is determined by the eddy-current demagnetizing effect accompanying the motion of the wall. The wall moves at such a rate that the demagnetizing field strength is equal to the difference between the applied field strength and the basic field strength corresponding to the wall position.

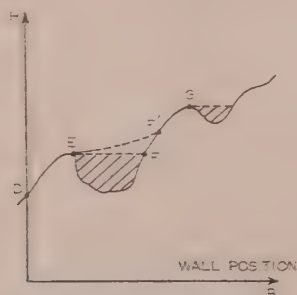


Fig. 8.—Diagram illustrating stability of a domain boundary wall.

This is best understood with reference to Fig. 8. If the field strength H is gradually increased through the range shown in Fig. 8, the wall will move in a stable manner first from D to E and then at E the motion will become unstable, unless H is suddenly reduced, and the wall will travel suddenly to F. As H is increased further the wall will again move in a stable manner to G, where a further sudden movement will occur.

The wall position may be regarded as a measure of flux density B . Thus it may be said that the shaded areas in Fig. 8 correspond to the B/H loss, or the hysteresis loss, that one can measure from the hysteresis loop.

Now, if it takes a finite time for the wall to travel from E to F, say, under alternating conditions of magnetization the value of H will have increased slightly by the time F is reached. We may expect F to be at some point F', in view of this. The motion of

the wall, expressed as a relationship between B and H , will then be as denoted by EF'. This corresponds to a greater loss and an increase of the speed of boundary movement under alternating conditions of magnetization.

However, this aspect of the increased loss due to finite speed of domain-wall movement can be accounted for by normal eddy-current action. Another aspect appears when we consider the effects of the time-lag near the corners of the hysteresis loop.

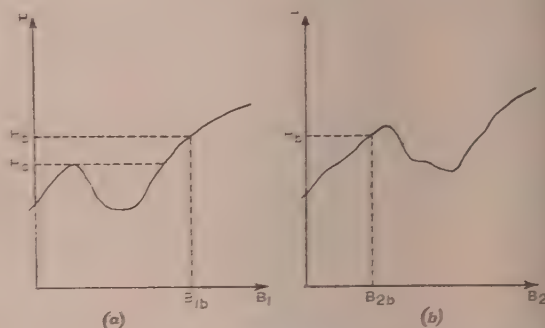


Fig. 9.—Diagram showing effect of magnetization involving more than one domain boundary wall.

In Figs. 9(a) and 9(b) two B/H relationships such as that shown in Fig. 8 are presented, and may be considered as applying respectively to each of two parallel flux paths through a specimen under consideration.

As H is increased slowly the value of B_1 as derived from Fig. 9(a) will experience a sudden rise when H reaches H_a . When H reaches H_b , let us assume that the maximum value of H has been reached. Then, the net flux will be $B_{1b} + B_{2b}$. Now consider H to be increased in a similar way but very quickly. When H has reached the maximum the value of B_2 will be B_{2b} , as before, but, owing to the delayed motion of the wall, B_1 will be appreciably less than B_{1b} . Thus, for a given range of magnetizing field the maximum flux has been reduced. When the wall has ceased to move in an unstable manner H has fallen, causing B_2 to have fallen below B_{2b} , and B_1 does not quite rise to B_{1b} . The result is that, to carry out a test using a similar range of flux, H has to be increased beyond H_b , and if it is so increased the hysteresis loss will be increased because other discontinuous wall-movements will occur. It follows that, owing to the time-lag associated with domain-boundary movements, hysteresis loss is increased and the dynamic hysteresis loss is therefore greater than the static hysteresis loss.

The phenomenon arises owing to a time-lag and the foregoing experimental work has been directed to the detection and measurement of this time-lag. It has been found to decrease with increasing frequency and to have a time-constant in ordinary steel of about 3×10^{-5} sec at a frequency of 300 c/s, falling to less than 10^{-5} sec at 10 kc/s.

It is of interest to compare this numerically with the time-lag we may expect to occur on the basis of the above-described mechanism.

A typical distance between domain walls in steel is 50 microns and a typical sudden jump of a domain wall is through a distance of 10 microns. Also, a typical coercive force is 2 oersteds. The coercive force is a convenient measure of the depth of a well such as that shown in the curve of Fig. 8. These data are sufficient for us to predict a time-constant associated with the movement of a domain boundary wall in steel.

In Fig. 10, XX, YY and ZZ represent three adjacent domain walls. They are separated by a distance D cm. Within the

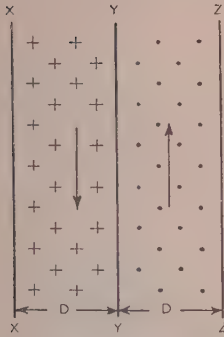


Fig. 10.—Diagram of a simple magnetic domain configuration showing local eddy-current action accompanying wall movement.

domains between the walls the permeability is effectively unity.* It is assumed that the domain between XX and YY is magnetized to a saturation flux density B_s in a direction which may be regarded as into the paper and the domain between YY and ZZ is similarly magnetized but in the opposite direction. It is assumed that owing to the application of a magnetizing field H_{coer} the wall YY suddenly moves to the right through a distance d cm. The induced e.m.f. per square centimetre of wall is $B_s \left(\frac{\partial x}{\partial t} \right) \times 10^{-8}$ where x is the displacement of the wall YY in centimetres. This e.m.f. will cause eddy currents to flow in all the surrounding material. It is presumed that this material is so extensive that eddy-current screening occurs which localizes the eddy-current effects to the system shown in Fig. 10. This is conveniently represented by assuming that the boundaries XX and ZZ move with the boundary YY to preserve their relative positions. No eddy-currents are, on this basis, induced in the material outside the system shown. Since the permeability within the domains is small, no eddy-current skin effects are likely to occur here and we may presume that a uniform current flows around the domains. This current is shown in Fig. 10 by the arrows. The resistance of the current path per square centimetre of boundary is $2\rho/D$ ohms. The value of the current $2B_s \left(\frac{\partial x}{\partial t} \right) \times 10^{-8} \frac{D}{2\rho}$. This produces a demagnetizing field which balances the coercive force H_{coer} .

$$\text{Thus } H_{coer} = \frac{4\pi}{10} B_s \left(\frac{\partial x}{\partial t} \right) \times 10^{-8} \frac{D}{\rho} \quad (10)$$

The value of $\partial x / \partial t$ is approximately d/τ and consequently τ may be calculated as

$$\tau = \frac{4\pi}{10^9} \frac{B_s D d}{\rho H_{coer}} \quad (11)$$

D , d , and H_{coer} are given the values shown above, and the values of ρ and B_s are 1.7×10^{-5} and 20000 respectively, it will be seen that τ has the value of 3.7×10^{-5} sec.

* Flux density B is given by the expression $4\pi M + H$, where M is the intensity of magnetization produced by the field H . In a saturated medium M has a fixed value M_s , and the value of $\partial B / \partial H$ becomes unity. It is this differential value of permeability which is effective when eddy-currents within the medium are concerned. Since a state of magnetization exists within a domain even in a zero field, $4\pi M_s$ may be regarded as a flux density appropriate to the state of magnetization and conveniently denoted B_s . However, an applied field which is not zero is necessary to promote boundary movement, and, in deriving a simple expression for an induced e.m.f. it is reasonable to assume that the magnetizing field is negligibly small in comparison with B_s .

This is of the appropriate order of magnitude to explain the observed time-lag effects.

When the hysteresis loop is displaced by a polarizing field the hysteresis loss is increased. The coercive force also increases. Thus we may expect τ to fall as the experiments indicate.

(6) CONCLUSIONS

By carrying out tests on thick steel specimens it is reasonable to assume that classical eddy-current theory applies, provided that the magnetic skin thickness is maintained appreciably greater than the dimensions of the magnetic domains in the steel. However, the above experiments have shown that there is a discrepancy between the actual and theoretical eddy-current effects in spite of this. This result is supported by the results of some special tests which the author⁵ has carried out on electrical sheet steels, proving that the eddy-current anomaly could not be wholly attributed to the inhomogeneity arising from magnetic domain configuration in low-loss electrical sheet. The discrepancy can be explained in terms of a time-lag—which has been measured and found to be frequency dependent, having a value ranging from about 3×10^{-5} sec at a few hundred cycles per second to a value of about 10^{-5} sec at 10 kc/s. A theory by which this time-lag can be explained has been proposed, and it has been shown that this theory gives a good qualitative and quantitative explanation of the time-lag effect.

(7) ACKNOWLEDGMENTS

The author acknowledges a sincere debt of gratitude for the awards of a Whitworth Senior Scholarship and The Institution I.M.E.A. Research Scholarship which enabled him to undertake this work. He is also grateful for the facilities and kind assistance afforded him at the Cambridge University Engineering Laboratories, and particularly the helpful encouragement shown him by Mr. G. E. Middleton.

(8) REFERENCES

- (1) BRAILS福德, F.: "Investigation of the Eddy-Current Anomaly in Electrical Sheet Steels," *Journal I.E.E.*, 1948, **95**, Part II, p. 38.
- (2) HEYS, B.: "Anomalous Eddy-Current Effects in Electrical Sheet Steels," Ph.D. Thesis, Cambridge, 1950.
- (3) STEWART, K. H.: "Losses in Electrical Sheet Steel," *Proceedings I.E.E.*, Paper No. 785 M, December, 1948 (**97**, Part II, p. 121).
- (4) LANDAU, L., and LIFSHITZ, E.: "Theory of the Dispersion of Magnetic Permeability in Ferromagnetic Bodies," *Physikalische Zeitschrift der Sowjetunion*, 1935, **8**, p. 153.
- (5) ASPDEN, H.: "The Eddy-Current Anomaly in Electrical Sheet Steel" (see page 272).
- (6) LATOUR, M.: "Notes on Losses in Sheet Iron at Radio Frequencies," *Proceedings of the Institute of Radio Engineers*, 1919, **7**, p. 61.
- (7) BETHENOD, J.: "Iron Losses in Sheets at High Frequency," *La Lumière Electrique*, 1916, **34**, 2nd series, p. 73.
- (8) THOMSON, J. J.: "On the Heat Produced by Eddy-Currents in an Iron Plate exposed to an Alternating Magnetic Field," *Electrician*, 1892, **28**, p. 599.
- (9) ASPDEN, H.: "Eddy-Currents in Solid Cylindrical Cores having Non-Uniform Permeability," *Journal of Applied Physics*, 1952, **23**, p. 523.

NOTE ON THE FOURIER COEFFICIENTS FOR CHEBYSHEV PATTERNS

By Dr. H. E. SALZER.

(The paper was first received 26th August, and in revised form 23rd November, 1955. It was published as an INSTITUTION MONOGRAPH in February, 1956.)

SUMMARY

The main purpose of the paper is to obtain a single formula for calculating feeding coefficients for optimum beam patterns. The new formula derived is equivalent to a set of formulae given by DuHamel, but is much simpler in form and more convenient for computation. It is applicable to Riblet's generalization of Dolph's procedure for finding optimum beam patterns in the design of linear arrays, for an odd number of elements when the distance between sources is less than half the wavelength. Whereas DuHamel's formulae involve a number of separate summation expressions, each of which has terms involving considerable multiplication and division, the corresponding terms in the present, more concise, expression have simpler integral coefficients involving the product of only two binomial coefficients. The mathematical problem solved is to find the explicit formula for b_m in the expression of $T_n(ax + b)$ in the form $\sum_{m=0}^n b_m T_m(x)$, where $T_m(x) = \cos(m \arccos x)$ for $m > 0$ and $T_0(x) = \frac{1}{2}$.

(1) INTRODUCTION

The use of Chebyshev polynomials in the design of linear arrays to produce optimum beam patterns has been developed by Dolph¹ and Riblet² for broadside arrays. They show how to obtain the explicit expressions for the feeding coefficients which will synthesize a pattern for which the beam width is a minimum for a given side-lobe level, or for which the side-lobe level is a minimum for a given beam width. Dolph found the optimum pattern for either an odd or an even number of elements when $d \geq \lambda/2$ (where d is the distance between sources), and Riblet generalized Dolph's procedure to find the optimum pattern for an odd number of elements when $d < \lambda/2$.

In the former (Dolph) case, where the number of elements is either $2n$ or $2n + 1$, the essential computational problem is the expression of the Chebyshev polynomial $T_n(z_0 x)$ as a series in terms of $T_m(x)$, where $T_m(x) = \cos(m \arccos x)$ for $m > 0$ and $T_0(x) = \frac{1}{2}$. A solution has been given by Barbieri³ and in somewhat simpler form by Salzer.⁴ For the general (Riblet) case, where the number of elements is $2n + 1$, the computational problem is more involved, since it requires the expression of $T_n(ax + b)$ in terms of $T_m(x)$, the Dolph case being merely the special case when $b = 0$. Formulae for the computation of the coefficients of $T_m(x)$ for the general Riblet case are given by DuHamel,⁵ who also determined the explicit expressions for the feeding coefficients in terms of a and b for $n = 2$ to $n = 8$. However, the DuHamel formulae, which involve four separate summation expressions (depending upon the parity relation between n and m), have elements in the summation of the form B_s , where the B_s are again expressed by four separate summation expressions depending upon the parity relation between r and s . Furthermore, each term in the summations for B_s is rather complicated for calculation, since it is presented as a fraction where in the numerator among other factors is a factorial multiplied by another integer and a power of 2, while the denominator is a product of three factorials. Although the coefficients of

$T_m(x)$ in this general case apparently must be expressed as double summations (unlike the simpler single-summation form in the Dolph case when $b = 0$), it is possible to derive formulae for computing them which are much simpler than those given by DuHamel. The purpose of the paper is to derive formulae equivalent to those in DuHamel for the feeding coefficients, but whose terms are given in simpler integral form and where, corresponding to the above-described computation of $2^{st}u!/o!p!q!$ as part of the work of obtaining DuHamel's B_s , one has here only to take the product of two binomial coefficients which can be found in standard tables,^{6,7} and in the literature for larger values of n (see Glaisher⁸).

(2) COMPUTATION

To express $T_n(ax + b)$ in the form $\sum_{m=0}^n b_m T_m(x)$ we begin with the substitution $z = ax + b$ in the general formula for $T_n(z)$ to obtain

$$T_n(ax + b) = \sum_{r=0}^{[n/2]} (-1)^r 2^{n-2r-1} \left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] (ax + b)^{n-2r} \quad (1)$$

where $\binom{m}{k}$ denotes the binomial coefficient $\frac{m!}{k!(m-k)!}$ and $[y]$ denotes the largest integer not exceeding y . By the binomial theorem,

$$(ax + b)^{n-2r} = \sum_{k=0}^{n-2r} \binom{n-2r}{k} a^k b^{n-2r-k} x^k \quad (2)$$

and x^k is expressible in terms of $T_m(x)$ by the formula

$$x^k = \frac{1}{2^{k-1}} \sum_{j=0}^{[k/2]} \binom{k}{j} T_{k-2j}(x) \quad (3)$$

Thus, eqns. (1), (2) and (3) may be combined to give

$$T_n(ax + b) = \sum_{r=0}^{[n/2]} (-1)^r 2^{n-2r} \left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] \sum_{k=0}^{n-2r} \binom{n-2r}{k} \left(\frac{a}{2} \right)^k b^{n-2r-k} \sum_{j=0}^{[k/2]} \binom{k}{j} T_{k-2j}(x) \quad (4)$$

In eqn. (4), add together all coefficients of $T_{k-2j}(x)$ for which $k - 2j$ has the same value m . For each particular value of r , m will range from 0 to $n - 2r$. The range of j and k is shown by the dark dots in the lattice diagram, Fig. 1.

For any value of m , Fig. 1 shows how $m = k - 2j$ begins with $k = m, j = 0$, then $k = m + 2, j = 1$, next $k = m + 4, j = 2$, etc. Now when m and $n - 2r$, or m and n , have the same parity, the uppermost dark dot on the line $m = m$ is at $k = n - 2r, j = (n - 2r - m)/2$; when m and $n - 2r$, or m and n , have different parity, the uppermost dark dot on the line $m = m$ is at $k = n - 2r - 1, j = (n - 2r - m - 1)/2$. But the

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Salzer is at the Diamond Ordnance Fuze Laboratories, Washington, D.C.

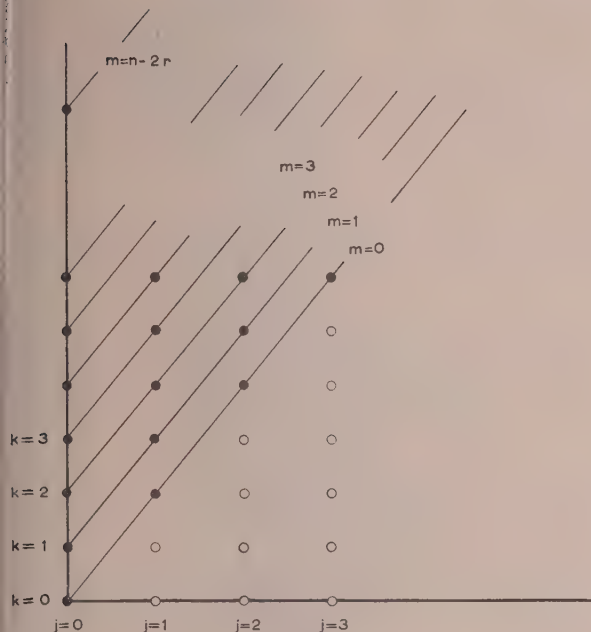


Fig. 1

value of the uppermost dark dot for both situations can be expressed by $k = m + 2\left\{[(n-m)/2] - r\right\}$, $j = [(n-m)/2] - r$, thus the double summation with respect to j and k for each value of r in eqn. (4) can be changed to a double summation with respect to j and m , to give

$$T_n(ax + b) = \sum_{r=0}^{[n/2]} (-1)^r 2^{n-2r} \left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] \sum_{m=0}^{n-2r} \sum_{j=0}^{[(n-m)/2]-r} \binom{m+2j}{j} \left(\frac{n-2r}{m+2j} \right) \left(\frac{a}{2} \right)^{m+2j} b^{n-2r-m-2j} T_m(x) \quad (5)$$

Now the operation of $\sum_{r=0}^{[n/2]} \sum_{m=0}^{n-2r}$ in eqn. (5) is seen to be for m and r corresponding to the dark dots in Fig. 2. The operation $\sum_{r=0}^{[n/2]} \sum_{m=0}^{n-2r}$ implies summation for every value of r , over each of the horizontal layers of points as far as $n - 2r$. Interchanging the orders of summation, for every m ranging from $m = 0$ to

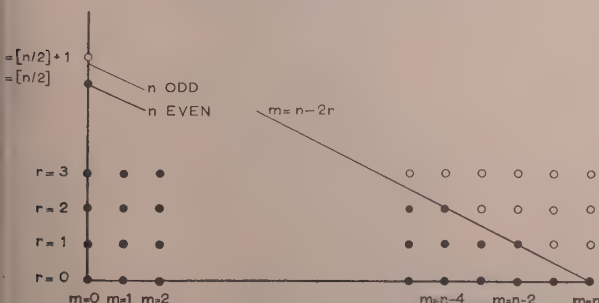


Fig. 2

$m = n$, each of the vertical layers of points is summed from $r = 0$ as far as $r = [(n-m)/2]$, so that $\sum_{r=0}^{[n/2]} \sum_{m=0}^{n-2r} \equiv \sum_{m=0}^n \sum_{r=0}^{[(n-m)/2]}$.

Then eqn. (5) can be expressed in the form

$$T_n(ax + b) = \sum_{m=0}^n \sum_{r=0}^{[(n-m)/2]} (-1)^r 2^{n-2r} \left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] \sum_{j=0}^{[(n-m)/2]-r} \binom{m+2j}{j} \left(\frac{n-2r}{m+2j} \right) \left(\frac{a}{2} \right)^{m+2j} b^{n-2r-m-2j} T_m(x) \quad (6)$$

or

$$T_n(ax + b) = \sum_{m=0}^n b_m T_m(x) \quad (7)$$

where, from eqn. (6) and after slight rearrangement of the powers of a , b and 2 , the coefficient b_m is expressed as

$$b_m = (2b)^{n-m} a^m \sum_{r=0}^{[(n-m)/2]} \left(\frac{-1}{4b^2} \right)^r \left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] \sum_{j=0}^{[(n-m)/2]-r} \binom{m+2j}{j} \binom{n-2r}{m+2j} \left(\frac{a^2}{4b^2} \right)^j \quad (8)$$

(3) CONCLUSION

Eqn. (8) seems to be the simplest explicit expression for the numerical computation of b_m , since for the various special values of a and b that are usually determined only approximately from physical considerations of desired side-lobe level and distance between sources, it is as easy to compute powers of $-1/4b^2$ and $a^2/4b^2$ as powers of a and b . When the coefficients

$$\left[2 \binom{n-r}{r} - \binom{n-r-1}{r} \right] \binom{m+2j}{j} \binom{n-2r}{m+2j}$$

have been calculated, it is easy to obtain the explicit coefficients of the various terms in $a^{m+2j} b^{n-m-2r-2j}$ by multiplication by the integer $(-1)^r 2^{n-m-2r-2j}$ (which is the form in which DuHamel tabulates b_m). Some computers might prefer to

replace the $2 \binom{n-r}{r} - \binom{n-r-1}{r}$ in eqn. (8) by $\frac{n}{r} \binom{n-r-1}{r-1}$

for $r > 1$, because the n may be taken outside the double summation in eqn. (8), leaving only a single operation of division.

(4) REFERENCES

- (1) DOLPH, C. L.: "A Current Distribution for Broadside Arrays which Optimizes the Relationship between Beam Width and Side Lobe Level," *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 335.
- (2) RIBLET, H. J.: Discussion on Reference 1, *ibid.*, 1947, **35**, p. 489.
- (3) BARBIERE, D.: "A Method for Calculating the Current Distribution of Tschebyscheff Arrays," *ibid.*, 1952, **40**, p. 78.
- (4) SALZER, H. E.: Discussion on Reference 3, *ibid.*, 1954, **42**, p. 1021.
- (5) DUHAMEL, R. H.: "Optimum Patterns for Endfire Arrays," *ibid.*, 1953, **41**, p. 652.

- (6) "Mathematical Tables," from the "Handbook of Chemistry and Physics" (Chemical Rubber Publishing Co., Cleveland, 1947), Eighth edition, p. 250.
- (7) BRITISH ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE: "Mathematical Tables: Vol. IX: Table of Powers" (University Press, Cambridge, 1940), p. 132.
- (8) GLAISHER, J. W. L.: "Table of Binomial-Theorem Coefficients," *Messenger of Mathematics*, 1917, **47**, p. 97.
- (9) NATIONAL BUREAU OF STANDARDS: "Tables of Chebyshev Polynomials $S_n(x)$ and $C_n(x)$," *Applied Mathematical Series*, 1952, **9** [see formulae (35) and (38)].
-

RADIATION PATTERNS OF CIRCUMFERENTIAL SLOTS ON MODERATELY LARGE CONDUCTING CYLINDERS

By J. R. WAIT, M.A.Sc., Ph.D., and J. KATES, M.A., Ph.D.

(The paper was first received 31st August, and in revised form 21st November, 1955. It was published as an INSTITUTION MONOGRAPH in February, 1956.)

SUMMARY

Computed patterns are presented for thin half-wave circumferential slots on circular conducting cylinders of infinite length. The cases considered are for single and diametrically opposed slots on cylinders whose circumferences vary from 3 to 21 wavelengths.

(1) INTRODUCTION

The radiation from slots on circular cylinders of infinite length and perfect conductivity has been discussed on several occasions.^{1,2,3,4,5} Attention has been paid mainly to the cases in which the diameter $2a$ of the cylinder is not much greater than the wavelength λ . For large cylinders the zonal harmonics solution involved converges very poorly. The numerical work then becomes so extensive that it should be programmed for automatic computers.

It is the purpose of the paper to report the results of some further extensive computations undertaken to determine the theoretical patterns of thin half-wave circumferential slots cut in circular cylinders whose circumferences vary from 3 to 21λ . The infinite-series expressions arising in the solution were programmed on the Ferut computer at the University of Toronto. As a check on the final results obtained, an alternative formulation of the problem is carried out, and leads to a more rapidly convergent series for certain cases.

(2) THE ZONAL HARMONIC SOLUTION

With respect to a cylindrical co-ordinate system (ρ, ϕ, z) the slot at $z = 0$ extends from $\phi = \pi/2 - \lambda/4a$ to $\pi/2 + \lambda/4a$ along the circumference of the cylinder defined by $\rho = a$, as shown in Fig. 1. Since the slot is assumed to be thin, the voltage distribution along it is sinusoidal. The radiation field in the plane $\theta = 0$ is plane polarized with the electric vector in the axial direction and is given by⁵

$$E_z = -\frac{\varepsilon}{\rho} \frac{jk\rho}{\rho} VP(\phi)\varepsilon^{j\omega t} \quad (1)$$

where V is the effective voltage at the centre of the slot, $k = 2\pi/\lambda$, and ω is the angular frequency. The pattern function $P(\phi)$ is

$$P(\phi) = \sum_{p=0}^{\infty} \epsilon_p \pi^2 \frac{x \varepsilon^{j(p-1)\pi/2}}{H_p^{(2)}(x)} \frac{\cos(p\pi/2x) \cos p(\phi - \pi/2)}{(x^2 - p^2)} \quad (2)$$

where $\epsilon_0 = 1$, $\epsilon_p = 2$ for $p \neq 0$, $x = ka = \frac{2\pi a}{\lambda}$ and $H_p^{(2)}(x)$ is the Hankel function of the second type of order p .

The results can be extended⁵ easily to the case of a pair of slots opposed diametrically on the surface and centred at $\phi = \pi/2$ and $\phi = -\pi/2$, as indicated in Fig. 1. The pattern function is then

$$P^+(\phi) = P(\phi) + P(\phi + \pi) \quad (3)$$

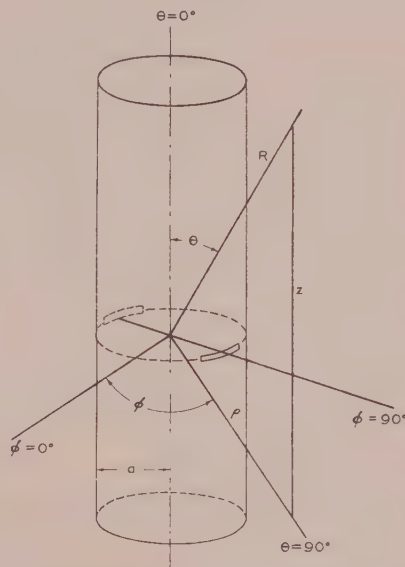


Fig. 1.—The co-ordinate system (ρ, ϕ, z) and the circular conducting cylinder of infinite length are illustrated. The positions of the circumferential slots are also indicated.

when the slots are fed in phase and

$$P^-(\phi) = P(\phi) - P(\phi + \pi) \quad (4)$$

when the slots are fed in anti-phase. The equatorial pattern $P(\phi)$ of the single slot centred at $\phi = 90^\circ$ is shown plotted in Fig. 2 for azimuthal angles ϕ from -90° to $+90^\circ$. It should be noted that $|P(\phi)|$ is an even function about $\phi = 90^\circ$. The values of $x = ka$ are indicated on the curves. To eliminate overlapping of the curves, the ordinate is shifted for each curve by a constant amount. The patterns $|P^+(\phi)|$ and $|P^-(\phi)|$ for the dual slots are shown in Figs. 3 and 4 respectively for ϕ from 0° to 90° . In this case $|P^+(\phi)|$ and $|P^-(\phi)|$ are even functions about both $\phi = 0^\circ$ and $\phi = 90^\circ$.

The data for the curves in Figs. 2 to 4 were obtained from the complex values⁷ of $P^+(\phi)$ and $P^-(\phi)$ tabulated in the Appendix.

It is interesting to note that the single-slot pattern is smoothly varying for directions behind the cylinder. This behaviour contrasts markedly with the corresponding equatorial plane patterns for an axial slot where numerous rear lobes are present.^{3,4} The structure of the pattern for the dual circumferential slots indicates that considerable interference takes place between the fields of each slot. In the case of the anti-phase connection, the pattern is identically zero at $\phi = 0^\circ$, owing to complete cancellation of the two individual fields.

A check on these patterns is to compare them with certain special cases that have been published elsewhere. For example,

Dr. Wait, who is at the National Bureau of Standards, Colorado, United States, was formerly with the Defence Research Board, Ottawa, Ontario.
Dr. Kates is at the Computation Centre, University of Toronto.

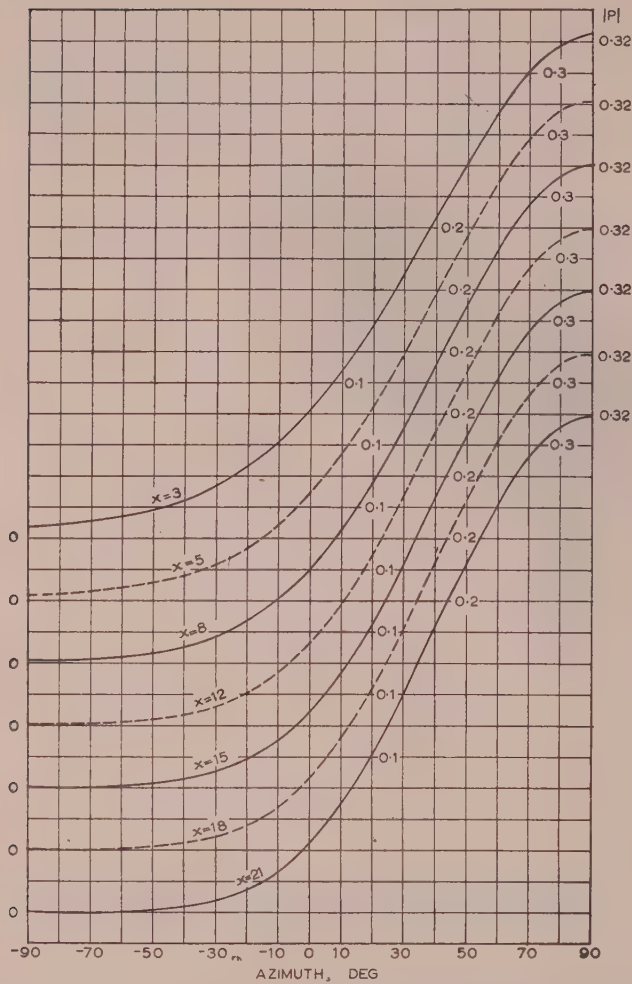


Fig. 2.—Equatorial H-plane patterns for single half-wave circumferential slot centred at $\phi = 90^\circ$ for various values, x , of the circumference in wavelengths.

Note.—Vertical scale is shifted for each curve.

Baillin⁶ has given extensive results for $ka = 8$ and 12 for single slots. Wait and Kahana⁵ have published both curves and tables for $ka = 2, 3$, and 5 for both single and dual circumferential slots. The agreement with the present results is good.

(3) AN ALTERNATIVE SOLUTION

It has been suggested by Sensiper,⁴ Imai⁸ and possibly others that the zonal harmonic-series solution for cylinder problems can be transformed to a more convergent series for large values of ka . Franz and Depperman³ have formulated the problem of diffraction by a cylinder in terms of an integral equation, which they solved approximately to obtain a solution similar to that of Imai. Actually, the method of Sensiper⁴ for the infinite axial slot on the circular cylinder is closely akin to the one employed by Watson¹⁰ for dipoles in the vicinity of large spheres. By a transformation with the aid of an integral in the complex plane, Watson succeeded in transforming the rigorous series of spherical zonal harmonics into a new highly convergent series. This technique has been discussed in detail by Bremmer.¹¹

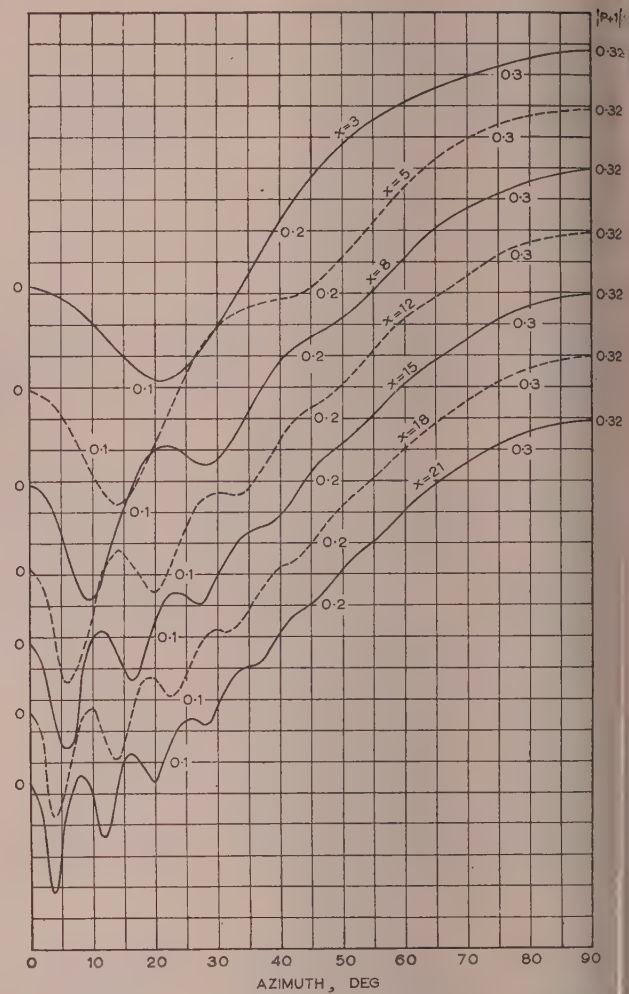


Fig. 3.—Equatorial H-plane patterns for diametrically opposed circumferential slots centred at $\phi = 90^\circ$ and -90° fed in phase for various values of x .

Note.—Vertical scale is shifted for each curve.

In this Section, the zonal harmonic series for the circumferential half-wave slot is treated by the above-mentioned method. The series solution is now written in the form

$$P(\phi) = \frac{x}{\pi^2 j} \sum_{p=-\infty}^{+\infty} \frac{e^{jp\pi/2} e^{-jp(\phi-\pi/2)} \cos(p\pi/2x)}{H_p^{(2)}(x)(x^2 - p^2)} \quad (1)$$

which is equivalent to the contour integral

$$P(\phi) = \frac{x}{2\pi^2} \int_{C_0} \frac{e^{j\nu\pi/2} e^{-j\nu(\phi-\pi/2)} \cos(\nu\pi/2x)}{H_\nu^{(2)}(x)(x^2 - \nu^2)} \left(\frac{e^{j\nu\pi}}{\sin \nu\pi} \right) d\nu \quad (2)$$

where the contour C_0 is shown in Fig. 5.

The contour C_0 is now deformed to the contour C which enclosed the zeros of $H_\nu^{(2)}(x)$, i.e. the poles of the integrand. These singular points lie in the second and fourth quadrants. It then follows that

$$P(\phi) = \frac{x}{2\pi^2} \times 2\pi j \sum \text{residues within } C \quad (3)$$

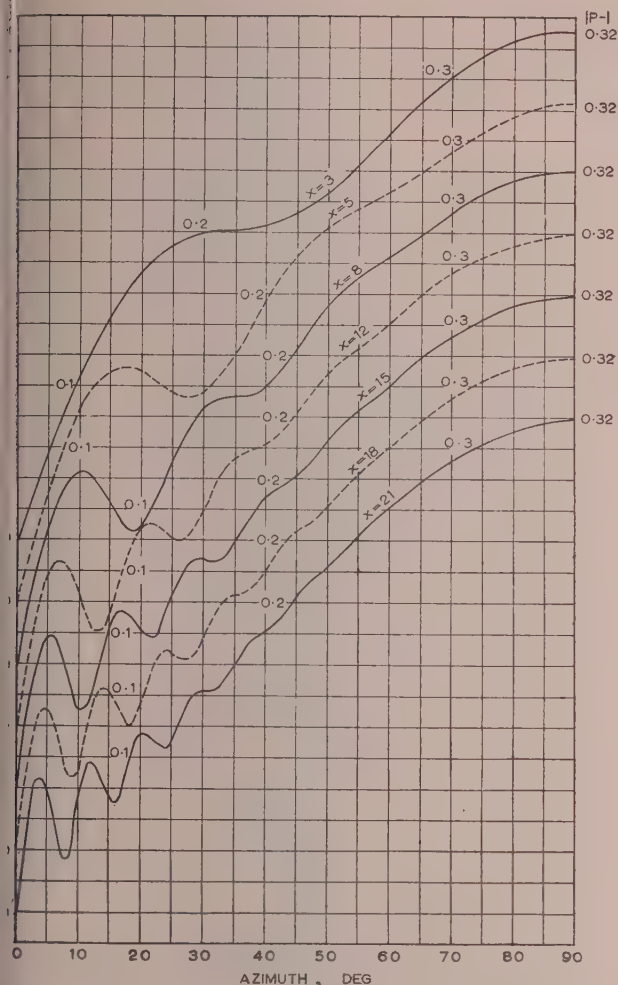


fig. 4.—Equatorial H-plane patterns for diametrically opposed circumferential slots centred at $\phi = 90^\circ$ and -90° fed out of phase for various values of x .

Note.—Vertical scale is shifted for each curve.

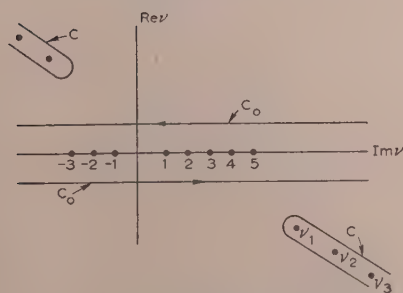


fig. 5.—The complex ν -plane showing the location of the zeros of $\sin \nu \pi$ and $H_\nu^{(2)}(x)$, and the contours C_0 and C .

which leads to

$$P(\phi) = x \left(\frac{2j}{\pi} \right) \sum_{m=1}^{\infty} \frac{\varepsilon^{j\nu_m \pi/2} \cos [\nu_m(\phi - 3\pi/2)] \cos (\nu_m \pi/2x)}{\left[\frac{\partial H_\nu^{(2)}(x)}{\partial \nu} \right]_{\nu=\nu_m} \sin \nu_m \pi} \quad (8)$$

where ν_m are the complex roots, with real parts greater than zero, of

$$H_\nu^{(2)}(x) = 0 \quad (9)$$

The determination of the roots ν_m from eqn. (9) has been discussed by Sensiper⁴ and Imai.⁸ Essentially the process is outlined as follows: The starting-point is Langer's¹² approximate representation for complex-order Hankel functions given by

$$H_\nu^{(2)}(z) = \left(\frac{\tan \alpha - \alpha}{\tan \alpha} \right)^{1/2} \varepsilon^{-j\pi/6} H_{1/3}^{(2)}[z(\sin \alpha - \alpha \cos \alpha)] \quad (10)$$

plus terms of order $\nu^{-5/4}$, where $\nu = z \cos \alpha$. Neglecting remainder terms, it then follows that

$$K_m = z(\sin \alpha - \alpha \cos \alpha) = r_m \varepsilon^{j\pi}$$

where K_m is the m th root of $H_{1/3}^{(2)}(K) = 0$ and r_m is the m th root of $J_{1/3}(r) + J_{-1/3}(r) = 0$. Now, for small α and large z , it follows that the desired roots α_m are given by

$$\alpha_m \approx \left(\frac{3K_m}{z} \right)^{1/3} = \delta_m \quad (11)$$

If we regard this as the first term of an expansion for α_m it can be written

$$\alpha_m = \sum_{n=1}^{\infty} a_n \delta_m^n \quad (12)$$

where the a_n coefficients are obtained in a straightforward manner by using

$$\sin \alpha - \alpha \cos \alpha = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{2n}{2n+1} \alpha^{2n+1} \quad (13)$$

along with eqns. (11) and (12), to yield

$$\alpha_m \approx \delta_m \left(1 + \frac{1}{30} \delta_m^2 + \frac{41}{12600} \delta_m^4 + \frac{97}{226800} \delta_m^6 + \dots \right) \quad (14)$$

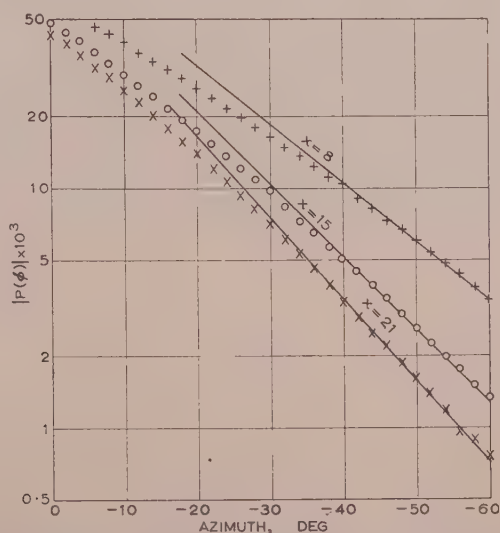


fig. 6.—The pattern functions computed from the residue series are shown by the solid curve; the points are calculated values from the rigorous zonal harmonic series.

and, since $\nu_m = z \cos \alpha_m$, a useful expansion for the roots is

$$\frac{\nu_m}{x} = 1 - \frac{1}{2}\delta_m^2 + \frac{1}{120}\delta_m^4 + \frac{1}{2800}\delta_m^6 + \frac{281}{9072000}\delta_m^8 + \dots \quad (15)$$

Using some well-known relations between Bessel functions, it can be shown that

$$\left. \frac{\partial H_{\nu}^{(2)}(x)}{\partial \nu} \right|_{\nu=\nu_m} \simeq -r_m^{1/2} e^{j\pi/3} H_{-2/3}^{(2)}(K_m) \frac{\alpha_m}{x^{1/2} [1 - (\nu_m/x)^2]^{1/4}} \quad (16)$$

which stems directly from eqn. (10).

Some special values of r_m and $H_{-2/3}^{(2)}(K_m)$ taken from available tables¹³ are:

$$\begin{array}{ll} r_1 = 2.38345 & H_{-2/3}^{(2)}(K_1) = 1.0390 \\ r_2 = 5.51020 & H_{-2/3}^{(2)}(K_2) = -0.68055 \\ r_3 = 8.64736 & H_{-2/3}^{(2)}(K_3) = 0.54291 \end{array}$$

The residue series solution for $P(\phi)$ given by eqn. (5) is rapidly convergent for large values of x (greater than about 8) for directions away from the slot ($-180^\circ < \phi < 0$). This region can be called the "shadow zone," since the slot is partly in the geometrical shadow with respect to the observer. The convergence of the series becomes very poor when $180^\circ > \phi > 0^\circ$, which is in the "illuminated zone."

As a check on the computations for the zonal harmonic solution, values of $|P(\phi)|$ obtained from eqn. (8) were plotted in Fig. 6 for the region 0° to -60° for x values of 8, 15 and 21. The indicated points in Fig. 6 are computed from eqn. (2) and are seen to agree very closely, in the shadow zone, with the residue series. No doubt the agreement would be even better if more terms corresponding to the roots ν_2 , ν_3 , etc., were included.

(4) CONCLUSIONS

The patterns of the single circumferential half-wave slot in the equatorial plane are seen to be a smoothly varying function of the azimuthal angle. When the slot is in the geometrical shadow region, the field becomes very small: in fact the level of the field radiated from the slot on the shadow side of the cylinder varies approximately as the cube root of the wavelength/diameter ratio of cylinder. On the other hand, the pattern of the single slot on the illuminated side of the cylinder is somewhat similar to that of a half-wave slot on an infinitely flat plate.

When there are two diametrically opposed slots on the cylinder, the patterns are considerably modified. In this case there is no shadow region in the optical sense, although diffraction still plays a major role, since it determines the nature of the interference pattern of the radiation from the two individual slots. It would also be very straightforward to calculate the equatorial plane pattern for any array of circumferential half-wave slots

by making use of the tabulated values of the functions $P^-(\phi)$ and $P^+(\phi)$.

It is believed that slotted-cylinder aeriels of this type mounted in a vertical position can be very convenient as a source of a vertically polarized radiation in the equatorial plane.

(5) REFERENCES

- (1) SINCLAIR, G.: "The Patterns of Slotted Cylinder Antennas," *Proceedings of the Institute of Radio Engineers*, 1948, 36, p. 1487.
- (2) PAPAS, C. H.: "Radiation from a Transverse Slot in an Infinite Cylinder," *Journal of Mathematics*, 1950, 28, p. 227.
- (3) SILVER, C., and SAUNDERS, W. K.: "Radiation from a Transverse Slot in a Circular Cylinder," *Journal of Applied Physics*, 1950, 21, p. 749.
- (4) SENSIPER, S., STERNS, W. G., and TAYLOR, T. T.: "A Further Study of the Patterns of Single Slots on Circular Conducting Cylinders." Paper presented at Spring International Scientific Radio Union (U.R.S.I.) Meeting, April 24, 1951.
- (5) WAIT, J. R., and KAHANA, S.: "Calculated Patterns of Circumferential Slots on a Circular Conducting Cylinder," *Canadian Journal of Technology*, 1955, 33, p. 77.
- (6) BAILLIN, L. L.: "The Radiation Field produced by a Slot in Large Circular Cylinder," *Transactions of Professional Group on Antennas and Propagation* (Institute of Radio Engineers), 1955, 3, p. 128.
- (7) The tabulated values of $P_+(\phi)$ and $P_-(\phi)$ were obtained from the Ferut computer and they are accurate to about four significant figures.
- (8) ISAO IMAI: "Diffraction of Electromagnetic Waves on a Circular Cylinder," *Zeitschrift für technische Physik*, 1954, 137, p. 31.
- (9) FRANZ, W., and DEPPERMAN, K.: "Theory of the Creeping Wave in Diffraction by a Cylinder," *Annalen der Physik*, 1952, 10, p. 361.
- (10) WATSON, G. N.: "The Diffraction of Radio Waves by the Earth," *Proceedings of the Royal Society*, 1919, 95, p. 546.
- (11) BREMMER, H.: "Terrestrial Radio Waves, Theory of Propagation" (Elsevier Publishing Co., Amsterdam, 1949) chapter 3.
- (12) LANGER, R. E.: "On the Asymptotic Solutions of Differential Equations with Application to the Bessel Functions of Large Complex Order," *Transactions of the American Mathematical Society*, 1932, 34, p. 447.

(6) APPENDIX

(6.1) Values of the amplitude and phase of $P_+(\phi)$ and $P_-(\phi)$ are given in Tables 1-7 for values of ϕ from 0° to 90° .

Table 1

 $ka = 3$

Azimuth, ϕ	P_+		P_-	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.1644	250.73	0.0000	0.00
2	0.1633	251.08	0.0220	316.94
4	0.1602	252.12	0.0437	317.19
6	0.1551	253.92	0.0648	317.61
8	0.1484	256.59	0.0849	318.20
10	0.1405	260.25	0.1038	318.98
12	0.1317	265.11	0.1212	319.94
14	0.1229	271.39	0.1370	321.11
16	0.1149	279.30	0.1510	322.50
18	0.1087	288.89	0.1631	324.13
20	0.1054	299.92	0.1733	326.01
22	0.1057	311.69	0.1816	328.17
24	0.1100	323.29	0.1881	330.63
26	0.1178	333.88	0.1930	333.42
28	0.1283	343.09	0.1964	336.54
30	0.1407	350.89	0.1986	340.03
32	0.1542	357.46	0.2000	343.89
34	0.1681	3.04	0.2009	348.11
36	0.1819	7.87	0.2016	352.67
38	0.1952	12.14	0.2024	357.54
40	0.2078	15.98	0.2038	2.65
42	0.2195	19.51	0.2059	7.91
44	0.2302	22.82	0.2089	13.22
46	0.2398	25.96	0.2130	18.50
48	0.2485	29.00	0.2181	23.63
50	0.2561	31.95	0.2242	28.54
52	0.2629	34.83	0.2311	33.17
54	0.2688	37.67	0.2386	37.48
56	0.2741	40.47	0.2466	41.45
58	0.2787	43.22	0.2548	45.09
60	0.2829	45.92	0.2631	48.39
62	0.2867	48.55	0.2712	51.37
64	0.2902	51.10	0.2791	54.06
66	0.2935	53.56	0.2866	56.47
68	0.2966	55.89	0.2936	58.62
70	0.2995	58.09	0.3001	60.53
72	0.3023	60.14	0.3060	62.23
74	0.3049	62.01	0.3113	63.71
76	0.3073	63.70	0.3159	64.99
78	0.3095	65.19	0.3199	66.09
80	0.3114	66.46	0.3233	67.01
82	0.3131	67.51	0.3260	67.75
84	0.3144	68.34	0.3281	68.33
86	0.3154	68.93	0.3296	68.73
88	0.3160	69.29	0.3305	68.98
90	0.3162	69.41	0.3308	69.06

Table 2

 $ka = 5$

Azimuth, ϕ	P_+		P_-	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.1378	247.34	0.0000	0.00
2	0.1352	248.03	0.0290	319.94
4	0.1276	250.22	0.0569	320.51
6	0.1158	254.29	0.0824	321.47
8	0.1009	261.04	0.1048	322.87
10	0.0850	271.97	0.1231	324.76
12	0.0716	289.16	0.1370	327.23
14	0.0658	313.01	0.1461	330.39
16	0.0713	337.80	0.1506	334.39
18	0.0857	357.02	0.1509	339.44
20	0.1044	10.36	0.1479	345.76
22	0.1239	19.97	0.1429	353.63
24	0.1420	27.51	0.1375	3.26
26	0.1576	34.00	0.1334	14.61
28	0.1703	40.04	0.1325	27.27
30	0.1798	46.01	0.1358	40.35
32	0.1864	52.17	0.1433	52.87
34	0.1906	58.69	0.1542	64.17
36	0.1930	65.70	0.1673	74.07
38	0.1946	73.21	0.1811	82.69
40	0.1961	81.19	0.1946	90.31
42	0.1985	89.51	0.2071	97.18
44	0.2021	97.97	0.2182	103.55
46	0.2075	106.34	0.2277	109.60
48	0.2144	114.39	0.2356	115.45
50	0.2227	121.97	0.2421	121.19
52	0.2319	129.00	0.2476	126.88
54	0.2416	135.45	0.2524	132.52
56	0.2512	141.36	0.2568	138.10
58	0.2605	146.75	0.2611	143.59
60	0.2692	151.70	0.2656	148.92
62	0.2771	156.26	0.2703	154.04
64	0.2840	160.47	0.2754	158.89
66	0.2901	164.37	0.2807	163.43
68	0.2954	167.99	0.2863	167.60
70	0.2999	171.33	0.2918	171.40
72	0.3037	174.40	0.2973	174.80
74	0.3070	177.20	0.3025	177.81
76	0.3097	179.71	0.3073	180.42
78	0.3120	181.93	0.3116	182.66
80	0.3139	183.85	0.3154	184.53
82	0.3155	185.43	0.3185	186.04
84	0.3167	186.68	0.3209	187.21
86	0.3175	187.59	0.3227	188.03
88	0.3180	188.13	0.3238	188.52
90	0.3182	188.31	0.3241	188.69

Table 3

 $ka = 8$

Azimuth, ϕ	P_+		P_-	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.1183	245.16	0.0000	0.00
2	0.1127	246.45	0.0376	322.27
4	0.0969	250.94	0.0716	323.40
6	0.0737	261.22	0.0987	325.40
8	0.0503	285.51	0.1165	328.49
10	0.0440	333.13	0.1240	333.07
12	0.0635	9.91	0.1214	339.82
14	0.0907	27.55	0.1110	349.98
16	0.1147	38.33	0.0972	5.49
18	0.1317	47.14	0.0871	28.04
20	0.1406	55.98	0.0885	54.82
22	0.1419	66.05	0.1020	78.47
24	0.1380	78.29	0.1215	96.17
26	0.1326	93.37	0.1407	109.70
28	0.1301	111.09	0.1561	121.17
30	0.1338	129.90	0.1662	132.02
32	0.1440	147.65	0.1712	143.18
34	0.1583	163.17	0.1728	155.23
36	0.1735	176.53	0.1732	168.40
38	0.1870	188.39	0.1748	182.50
40	0.1977	199.43	0.1795	196.92
42	0.2054	210.19	0.1875	210.90
44	0.2107	221.02	0.1981	223.92
46	0.2147	232.06	0.2099	235.79
48	0.2187	243.26	0.2215	246.63
50	0.2237	254.44	0.2318	256.66
52	0.2301	265.33	0.2405	266.12
54	0.2379	275.69	0.2475	275.22
56	0.2467	285.35	0.2534	284.05
58	0.2557	294.27	0.2586	292.66
60	0.2643	302.48	0.2636	301.00
62	0.2722	310.05	0.2688	309.01
64	0.2792	317.06	0.2744	316.57
66	0.2852	323.58	0.2803	323.62
68	0.2903	329.64	0.2863	330.09
70	0.2948	335.26	0.2922	335.95
72	0.2988	340.44	0.2978	341.20
74	0.3025	345.16	0.3028	345.84
76	0.3060	349.38	0.3072	349.90
78	0.3092	353.09	0.3108	353.39
80	0.3122	356.25	0.3137	356.33
82	0.3148	358.84	0.3160	358.74
84	0.3169	0.87	0.3176	0.61
86	0.3185	2.32	0.3187	1.94
88	0.3195	3.19	0.3194	2.75
90	0.3199	3.48	0.3196	3.01

Table 4

 $ka = 12$

Azimuth, ϕ	P_+		P_-	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.1039	243.80	0.0000	0.00
2	0.0933	246.08	0.0470	324.13
4	0.0648	255.79	0.0843	326.19
6	0.0324	295.46	0.1044	330.16
8	0.0446	12.97	0.1044	337.27
10	0.0806	37.28	0.0870	350.75
12	0.1067	49.35	0.0644	19.24
14	0.1160	60.76	0.0619	66.67
16	0.1094	75.77	0.0858	101.82
18	0.0953	98.68	0.1122	121.85
20	0.0891	131.33	0.1279	137.35
22	0.1010	164.15	0.1306	153.51
24	0.1225	188.67	0.1246	173.48
26	0.1409	207.53	0.1195	198.63
28	0.1504	224.82	0.1242	226.19
30	0.1517	243.24	0.1389	251.04
32	0.1501	264.16	0.1560	271.72
34	0.1522	287.08	0.1692	289.89
36	0.1610	309.78	0.1763	307.50
38	0.1745	330.39	0.1792	325.82
40	0.1880	348.77	0.1817	345.18
42	0.1986	5.78	0.1872	4.91
44	0.2054	22.36	0.1966	32.91
46	0.2101	39.05	0.2081	41.50
48	0.2149	55.87	0.2193	57.68
50	0.2215	72.45	0.2285	72.87
52	0.2301	88.28	0.2356	87.48
54	0.2396	103.05	0.2413	101.77
56	0.2489	116.76	0.2469	115.73
58	0.2571	129.57	0.2532	129.20
60	0.2639	141.67	0.2603	141.93
62	0.2697	153.16	0.2680	153.79
64	0.2750	164.09	0.2755	164.72
66	0.2804	174.38	0.2823	174.77
68	0.2859	183.93	0.2881	184.01
70	0.2915	192.68	0.2932	192.50
72	0.2969	200.56	0.2975	200.27
74	0.3019	207.58	0.3014	207.30
76	0.3062	213.75	0.3050	213.57
78	0.3098	219.09	0.3084	219.03
80	0.3127	223.63	0.3115	223.66
82	0.3149	227.36	0.3143	227.44
84	0.3165	230.28	0.3165	230.37
86	0.3176	232.39	0.3182	232.44
88	0.3182	233.66	0.3192	233.69
90	0.3185	234.08	0.3196	234.10

Table 5

 $ka = 15$

Azimuth, ϕ	P_{-}		P_{-}	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.0967	243.22	0.0000	0.00
2	0.0817	246.40	0.0529	325.12
4	0.0437	263.99	0.0892	328.05
6	0.0294	0.17	0.0982	334.22
8	0.0709	39.25	0.0800	347.72
10	0.0998	52.77	0.0515	24.37
12	0.1035	66.48	0.0581	88.04
14	0.0866	88.67	0.0912	120.17
16	0.0719	129.12	0.1127	138.87
18	0.0856	173.99	0.1134	158.04
20	0.1120	202.79	0.1015	184.86
22	0.1279	224.26	0.0972	221.73
24	0.1285	246.59	0.1122	257.29
26	0.1223	274.30	0.1332	284.22
28	0.1240	306.67	0.1458	306.95
30	0.1381	337.07	0.1478	330.38
32	0.1552	2.51	0.1465	356.93
34	0.1662	25.34	0.1512	25.43
36	0.1701	48.37	0.1636	52.35
38	0.1722	72.91	0.1780	76.23
40	0.1781	98.13	0.1883	98.24
42	0.1890	122.19	0.1940	119.97
44	0.2012	144.25	0.1982	142.16
46	0.2113	164.78	0.2045	164.41
48	0.2181	184.65	0.2138	185.78
50	0.2236	204.34	0.2244	205.68
52	0.2298	223.74	0.2339	224.26
54	0.2378	242.35	0.2414	241.94
56	0.2468	259.83	0.2475	259.03
58	0.2554	276.10	0.2534	275.56
60	0.2628	291.34	0.2601	291.32
62	0.2690	305.73	0.2675	306.10
64	0.2746	319.36	0.2748	319.77
66	0.2801	332.17	0.2815	332.37
68	0.2858	344.05	0.2873	343.98
70	0.2915	354.90	0.2923	354.67
72	0.2969	4.67	0.2968	4.45
74	0.3017	13.40	0.3010	13.29
76	0.3057	21.10	0.3049	21.13
78	0.3091	27.80	0.3086	27.93
80	0.3119	33.52	0.3118	33.67
82	0.3142	38.24	0.3144	38.35
84	0.3161	41.93	0.3164	41.96
86	0.3176	44.59	0.3177	44.54
88	0.3185	46.18	0.3185	46.09
90	0.3188	46.72	0.3188	46.60

Table 6

 $ka = 18$

Azimuth, ϕ	P_{+}		P_{-}	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.0912	242.82	0.0000	0.00
2	0.0714	247.07	0.0579	325.95
4	0.0259	283.32	0.0903	329.92
6	0.0481	32.44	0.0847	339.50
8	0.0883	51.07	0.0510	9.03
10	0.0960	65.61	0.0486	88.66
12	0.0744	91.78	0.0866	126.11
14	0.0613	147.56	0.1050	146.41
16	0.0878	195.46	0.0954	171.38
18	0.1122	222.28	0.0810	213.27
20	0.1137	247.60	0.0943	260.78
22	0.1032	281.96	0.1189	293.41
24	0.1067	324.29	0.1292	320.18
26	0.1268	0.13	0.1250	350.55
28	0.1423	28.82	0.1242	27.21
30	0.1452	57.28	0.1374	63.08
32	0.1443	89.39	0.1542	93.34
34	0.1517	123.05	0.1629	121.09
36	0.1662	153.75	0.1650	150.01
38	0.1784	181.30	0.1692	180.58
40	0.1845	208.13	0.1799	210.28
42	0.1885	235.69	0.1925	237.47
44	0.1957	263.31	0.2018	262.93
46	0.2065	289.47	0.2076	288.01
48	0.2171	313.79	0.2133	313.07
50	0.2251	336.92	0.2215	337.43
52	0.2311	359.49	0.2314	0.36
54	0.2376	21.53	0.2404	21.82
56	0.2456	42.58	0.2477	42.21
58	0.2541	62.31	0.2538	61.81
60	0.2620	80.73	0.2602	80.57
62	0.2685	98.03	0.2672	98.25
64	0.2743	114.36	0.2744	114.67
66	0.2800	129.69	0.2811	129.82
68	0.2858	143.89	0.2868	143.80
70	0.2916	156.85	0.2917	156.67
72	0.2969	168.55	0.2963	168.43
74	0.3015	179.03	0.3007	179.03
76	0.3053	188.31	0.3050	188.39
78	0.3086	196.42	0.3087	196.50
80	0.3114	203.31	0.3119	203.35
82	0.3139	208.97	0.3144	208.96
84	0.3160	213.37	0.3162	213.33
86	0.3176	216.50	0.3175	216.48
88	0.3185	218.37	0.3182	218.38
90	0.3188	219.00	0.3184	219.01

Table 7

 $ka = 21$

Azimuth, ϕ	P_+		P_-	
	Amplitude	Phase	Amplitude	Phase
deg		deg		deg
0	0.0868	242.52	0.0000	0.00
2	0.0618	248.09	0.0620	326.67
4	0.0174	332.56	0.0880	331.92
6	0.0660	44.90	0.0658	347.46
8	0.0916	60.03	0.0349	60.13
10	0.0736	83.33	0.0733	123.62
12	0.0524	145.96	0.0974	146.58
14	0.0830	202.71	0.0843	175.10
16	0.1054	231.26	0.0701	229.88
18	0.0972	262.87	0.0945	281.51
20	0.0881	311.68	0.1153	313.68
22	0.1075	359.28	0.1117	346.85
24	0.1267	33.52	0.1062	31.09
26	0.1272	67.22	0.1216	75.33
28	0.1245	107.83	0.1394	110.47
30	0.1369	149.09	0.1433	144.02
32	0.1532	184.22	0.1431	181.61
34	0.1598	217.16	0.1529	220.08
36	0.1616	252.13	0.1677	254.51
38	0.1697	287.90	0.1763	286.49
40	0.1827	321.06	0.1801	319.08
42	0.1926	351.81	0.1869	352.19
44	0.1981	22.17	0.1982	23.63
46	0.2044	52.63	0.2086	52.88
48	0.2139	81.98	0.2155	81.09
50	0.2241	109.47	0.2216	108.93
52	0.2320	135.55	0.2297	135.92
54	0.2383	160.81	0.2390	161.36
56	0.2453	185.23	0.2473	185.25
58	0.2533	208.35	0.2540	207.99
60	0.2613	229.96	0.2603	229.73
62	0.2682	250.19	0.2671	250.30
64	0.2741	269.23	0.2742	269.46
66	0.2799	287.08	0.2808	287.16
68	0.2858	303.60	0.2864	303.50
70	0.2916	318.68	0.2915	318.55
72	0.2967	332.31	0.2962	332.28
74	0.3012	344.55	0.3008	344.62
76	0.3050	355.42	0.3050	355.51
78	0.3084	4.90	0.3087	4.93
80	0.3115	12.95	0.3117	12.91
82	0.3141	19.54	0.3141	19.47
84	0.3161	24.65	0.3160	24.60
86	0.3175	28.29	0.3174	28.29
88	0.3183	30.47	0.3183	30.50
90	0.3185	31.19	0.3186	31.42

SOME PROPERTIES OF MAGNETRONS USING SPATIAL-HARMONIC OPERATION

By R. G. ROBERTSHAW and W. E. WILLSHAW, M.B.E., M.Sc.Tech., Member.

The paper was first received 28th April, and in revised form 6th October, 1955. It was published as an INSTITUTION MONOGRAPH in February, 1956.)

SUMMARY

To reduce the problems involved in the design of magnetrons of low voltage and power for operation at frequencies of the order of 1000 Mc/s and higher, a system is proposed in which the anode has only a few gaps, e.g. two or four. Operation at low anode voltage and magnetic field is achieved by electron interaction with the Fourier space components of the field.

A convenient and simple embodiment is described, and factors likely to result in operation different from the conventional multiple-circuit magnetrons are analysed briefly. In particular, the modes of oscillation and the magnitudes of the components of the different possible spatial harmonics are discussed for two types of anode circuit.

Details of operation of valves using both types of anode with two and four gaps are given, and comparisons are made where appropriate with the earlier analysis. The simple electrical structure of the two-gap anode enables a useful tuning range to be obtained by external-circuit adjustment, and some details of tuning performance are given.

The paper concludes with comments on noise performance and speed of build-up of oscillations.

LIST OF SYMBOLS

- N = Number of resonant circuits.
- ω = Angular frequency.
- V_{min} = Theoretical minimum voltages for oscillation.
- a = Anode radius.
- b = Cathode radius.
- r = A radius between a and b .
- e = Electron charge.
- m = Electron mass.
- n_0 = Fundamental mode number due to fields in N cavities.
- n = Component mode number.
- μ = An integer.
- θ = Angular position relative to "A" symmetry axis.
- θ_s = Angular width of slot.
- ϕ = Angular spacing between gaps 1 and 2 in m.s. anode.
- $K = 2\pi/2\phi$.
- E = Electric field in resonator gap.
- E_a = Total electric field at anode surface.
- E_n = Electric field of component n .
- V = R.F. voltage across gap.
- I_S = Anode current for start of oscillations.
- I_F = Anode current for finish of oscillations.
- Q_0 = Q-factor of unloaded valve anode circuit.
- Q_E = Q-factor of loaded valve anode circuits assuming $Q_0 = \infty$.
- Q_L = Q factor of loaded valve anode circuit.

(1) INTRODUCTION

The use of the multiple-circuit magnetron to generate a few watts of continuous-wave power at wavelengths of the order of 1 cm involves a design of valve which is difficult to construct where the voltage is less than 1000 volts, owing to the need both for small resonant circuits with narrow anode gaps and for an accurately central cathode. Such a valve having N resonant

circuits, where N may be in the range 18–24, is usually designed for operation in the π mode, i.e. with a phase difference of π between the voltages across successive gaps. Continuous interaction takes place between the rotating electron cloud and that rotating-wave component of field present at the anode having the velocity $\omega/2N$, which is the fundamental of the infinite series of waves into which the discontinuous field at the anode surface may be analysed.¹

The theoretical minimum voltage V_{min} at which oscillations may be generated—although the efficiency is zero—is given by equating the potential energy of an electron leaving the cathode with the kinetic energy of the electrons arriving at the anode, which are assumed to have the velocity of the field with which they are interacting. Thus

$$eV_{min} = \frac{1}{2}m(\omega a/2N)^2$$

In practice, a voltage about twice the value given by this expression needs to be used if good efficiency is to be obtained. It then follows that for fixed anode voltage, efficiency and anode radius the number of circuits must be increased in proportion to frequency. Further, for fixed anode voltage and frequency, change of number of circuits does not result in any change in the spacing between segments, or in segment gap-width. Finally, with increase in frequency and fixed number of segments and voltage, the gap width must decrease.

The need to increase the number of segments, and to reduce the gap width as the frequency is increased, introduces serious difficulties. Increase in number of segments and resonant circuits adds to the number of possible modes of resonance of the anode system and results in increased problems in ensuring satisfactory stability of operation in the required mode of resonance of the anode system. Reduction of gap width and segment width brings in mechanical difficulties of construction, together with problems of heat dissipation.

The question which has frequently been asked is whether these difficulties could be minimized by the use of Fourier space components of the travelling wave with which the electron stream interacts. Simple analysis shows that if the circuits of a multiple equal-cavity magnetron are excited at the frequency of one of the resonant modes, so that the pattern of the electromagnetic field within successive cavities is repeated n_0 times around the circumference, the electromagnetic wave within the interaction space between the anode surface and the cathode is composed of an infinite series of components. This arises from the discontinuous changes in electric field which take place at the anode surface owing to the presence of the metal segments of finite width. In an equal-cavity system these components have numbers $n_0 \pm \mu N$, μ being an integer. Putting $\mu = +1$ and $n_0 = \frac{1}{2}N$, corresponding to operation in the π mode, one component of this series is that having a mode number $n = 3N/2$, i.e. the same mode number as for a valve with $3N$ resonators instead of N . Thus, if the amplitude of the field of this component in the interaction space is sufficient, it should be possible to achieve operation by interaction with this component, with the advantage of the larger number of circuits but without the mechanical and electrical disadvantages.

Correspondence on Monographs is invited for consideration with a view to publication. The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

Experiments have been made to see whether operation of this type, requiring lower magnetic field and voltage, can be obtained, but so far as is known these have not been successful for valves having upwards of eight resonators with anode diameters in excess of 0.1λ . The lack of success in the cases so far tried may be ascribed to the rapid fall off of the amplitude of the desired component of the field away from the anode. Analysis shows that in the absence of space charge the amplitude of the tangential electromagnetic field at radius r is proportional to

$$(r/a)^{n-1}[1 - (b/a)^{2n}]$$

Thus, if n is large (it has a minimum value of 12 in the designs mentioned above) this field falls off very rapidly on leaving the anode surface. Consequently, it seems probable that in the cases so far investigated the magnitude of this field over the interaction space is too small to allow effective bunching of the electron stream to take place.

It was therefore decided to investigate the performance of valves particularly designed for this type of operation, where both the mode number of the wave component used and the anode radius were not too large, and the work described here is the outcome of this.

It was thought that if satisfactory operation could be obtained, it would lead to valves of relatively simple construction, and that, owing to the simpler resonance properties of the anode system, following the smaller number of resonant cavities used, it would be possible to achieve a useful tuning range by coupling the valve to a simple form of external-cavity tuner.

(2) DESCRIPTION OF EXPERIMENTAL VALVES

In the first place it was proposed to explore space-harmonic operation in an anode system which could be thought of as having 12 uniformly spaced resonators, eight of which had been omitted from the anode in two groups of four, and it was hoped to achieve operation by interaction between the electron stream and a component field in the interaction space having the same periodicity as that produced by the full number of resonators.

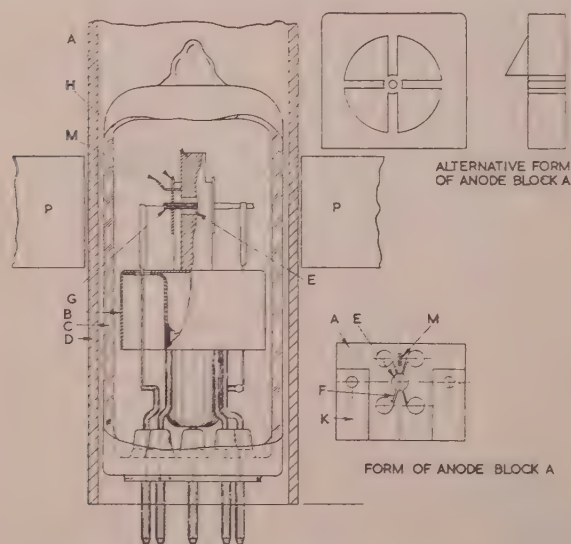


Fig. 1.—General arrangement of spatial-harmonic valve.

Fig. 1 shows the general form of valve used for these experiments. The anode system is formed in a thin block of copper, A, supported by a cup, B, within a glass bulb, C, fitting inside a

circular waveguide, D. The anode system consists of a central hole, E, of diameter 2 mm having four resonant circuits of the hole-and-slot type arranged around it. Adjacent slots, F, which are long and narrow have their centre lines on diameters displaced by an angle of 30° , i.e. they have the spacing of a system of 12 slots, although eight of these are missing. This system was called the missing-segment system (abbreviated m.s.). It will be clear from inspection of the Figure, in which the proportions of the holes and slots are arranged for an oscillating wavelength of 3 cm, that it would, in fact, be impossible to accommodate the full number of circuits in the space available. In an alternative design a further two adjacent resonant circuits were omitted leaving only two remaining. A tubular cathode, G, is supported centrally within the anode hole by mica cards, H, attached to the raised portions, K, of the copper block, A. Projecting from the end of this block from the tongue formed by two adjacent slots is a small wire, M, through which power is radiated from the oscillating system into the waveguide formed by the surrounding close-fitting metal tube, D.

The top surface of the cup, B, forms the zero-potential reference surface of the whole oscillatory system, and the cylindrical portion forms with the outer tube a choke to prevent transmission of power downwards to the base of the valve. The base carries pins to which connections are made to anode, cathode and heater wires, and which carry a support for the whole oscillatory system.

Operation takes place with the valve and its surrounding circular waveguide placed in a magnetic field of direction parallel to the cathode axis and formed between the poles, P, of a permanent magnet.

After a number of experimental valves of this type had been made it became clear that difficulties were arising in repeatability of operation, owing to inaccuracies in construction arising from an inability to machine accurately the long narrow slots used in the design. The use of a system having the same limited number of circuits (four or two) with gaps which are, however, symmetrically placed around the anode surface was then considered. Such a system permits considerable simplification in the method of forming the anode (e.g. hobbing may be used) and results in a valve of relatively cheap but accurate construction.

Fig. 1 shows an outline of the anode of the 4-segment version of this valve, which is referred to as the symmetrical-anode type (abbreviated s.a.). Since the width of adjacent gaps is no longer geometrically limited to a value such that the wave transit angle $n\theta_s$ at the anode surface is of the order of $\frac{1}{2}\pi$ (as it was with the m.s. system), the opportunity may be taken to increase this to a value close to π . This change reduces the amplitude of the space component for a given gap voltage (it is proportional to $\sin \frac{1}{2}n\theta_s / \frac{1}{2}n\theta_s$), and consequently the circuit load impedance must be increased in order to give the necessary gap-voltage increase.

(3) ANALYSIS OF MECHANISM OF OPERATION OF M.S. VALVE

(3.1) Modes of Resonance

The anode circuits, when not coupled to a load, may be considered as a first approximation to behave as the ring of lumped circuits. Such a system has four modes of oscillation, although formally two of these modes are different components of a doublet mode.

The first, the π mode, for which there is a 2λ standing wave of potential around the anode, is one in which currents in adjacent cavities differ in phase by π , as indicated in Fig. 2(a).

The second and third modes, the $\frac{1}{2}\pi$ modes, are those for which there is only a single standing wave of potential, as indicated in Figs. 2(b) and 2(c) for the two possible axes of symmetry A or E of a loosely coupled driving source. This driving source might

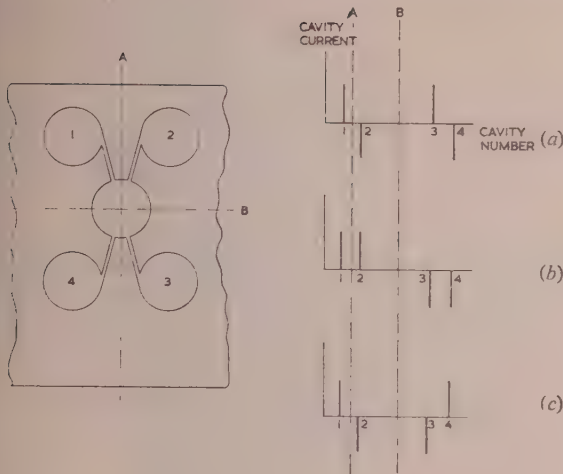


Fig. 2.—Distribution of current in m.s. anode.

- (a) π mode.
 (b) $\pi/2$ mode with A symmetry axis.
 (c) $\pi/2$ mode with B symmetry axis.

for example, be a short probe of the type shown in Fig. 1, suitably excited.

The fourth mode is the zero mode, for which currents in all four cavities are in phase, and there are consequently large potentials across the segments of anode surface joining slots 2 and 3 and slots 4 and 1. The frequency of this mode is very high, with the small ratio between anode diameter and wavelength used in these valves, so that it may be ignored so far as operation is concerned.

In later work it was decided to eliminate two further oscillatory circuits, leaving only circuits 1 and 2 coupled to an otherwise plain anode. Only two modes of resonance are then possible, the π mode and the zero mode of much higher frequency.

(3.2) Analysis of Interaction Field

(3.2.1) Oscillation in the π Mode (Four Segments).

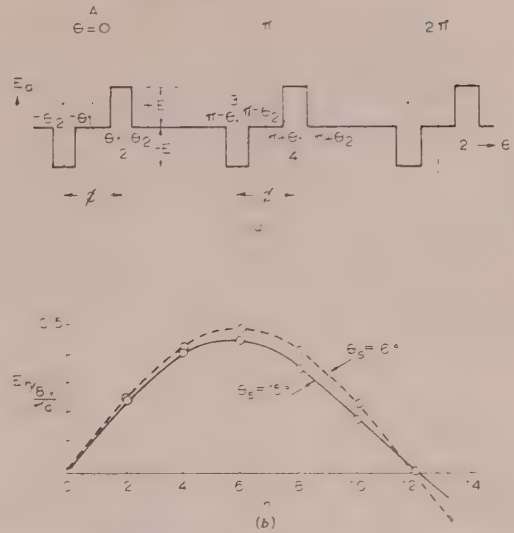
The total field present within the space between anode and cathode, arising from the fields present at the anode surface which are a consequence of the distribution of current in the resonant cavities discussed above, may be derived by Fourier analysis.

It can be assumed that, with the narrow gaps (in terms of oscillating wavelength) used in the design under discussion, and with the relatively large spacing between anode and cathode, the field at the anode is represented by the simple function shown in Fig. 3(a) when the anode is resonant in the π mode. Here the angular position $\theta = 0$ represents the axis of symmetry A of Fig. 2(a) and $\theta = \theta_1, \theta_2$ defines the position of the edges of the gap 2. The position of the edges of the gap 1 is defined by $\theta = -\theta_1, -\theta_2$. Gaps 3 and 4 are similarly placed round the line $\theta = \pi$. The spacing ϕ between the centre lines of gaps 1 and 2 and of 3 and 4 is set by the requirement $\phi \times 2K = 2\pi$, where K is an integer. Thus the spacing between near resonators is the same as it would be if $2K$ equally spaced resonators were used in the anode.

Fourier analysis of the field distribution having a peak value of $\pm E$ gives a total field at the anode surface

$$E_a = \frac{4E}{\pi} \sum_{n=-\infty}^{\infty} \frac{\cos n\theta_2 - \cos n\theta_1}{n}$$

with the limitation that only even values of n are valid.

Fig. 3.—Field distribution and component amplitudes for π mode (m.s.).

- (a) Field distribution.
 (b) Component amplitudes.

Inserting V , the r.f. voltage across the slots, the magnitude of the separate components is

$$E_n = \frac{8V}{\pi a} \left[\frac{\sin \frac{1}{2}n(\theta_2 - \theta_1) \sin \frac{1}{2}n(\theta_2 + \theta_1)}{n\theta_s} \right]$$

The variation of $E_n \pi a / 8V$ with n is shown in Fig. 3(b) for the case

$$\frac{1}{2}(\theta_2 - \theta_1) = \frac{1}{2}\theta_s = 7.5^\circ, \frac{1}{2}(\theta_2 + \theta_1) = 15^\circ$$

This characteristic shows a peak at $n = K = 6$, indicating that the component having the maximum amplitude is that which would also have maximum amplitude with $2K$ equally spaced gaps. The value of this maximum is

$$\frac{8V}{\pi a} \frac{\sin \frac{1}{2}n\theta_s}{n\theta_s}$$

The dotted curve shows the effect of reducing the slot width θ_s from 15° to 6° with the spacing of the slots unchanged.

(3.2.2) Oscillation in the $\pi/2$ Mode (Four Segments).

A similar analysis may be carried out for the two cases of oscillation in the $\pi/2$ mode, for which the field distributions are as shown in Figs. 4(a) and 4(b). The amplitudes of the separate components are now

$$E_{nA} = \frac{8V}{\pi a} \left[\frac{\sin \frac{1}{2}n(\theta_2 - \theta_1) \cos \frac{1}{2}n(\theta_2 + \theta_1)}{n\theta_s} \right]$$

for symmetry relative to the line A and

$$E_{nB} = \frac{8V}{\pi a} \left[\frac{\sin \frac{1}{2}n(\theta_2 - \theta_1) \sin \frac{1}{2}n(\theta_2 + \theta_1)}{n\theta_s} \right]$$

for symmetry relative to the line B. In both cases only odd values of n are permissible.

Values of the amplitudes of these components are shown in Fig. 4(c) for $\theta_s = 15^\circ$.

(3.2.3) Oscillation in the π Mode (Two Segments).

In the case of two segments, only the π mode is of interest, and for this the total field is just half that for the π mode with four segments, but now all values of n are present.

4.2.3) Oscillation in the π Mode (Two Segments).

The amplitude of the separate components is given by

$$E_n = \frac{4V}{\pi a} \left(\frac{\sin \frac{1}{2}n\theta}{n\theta_s} \right)$$

with the restriction that only odd values of n are possible. The envelope is accordingly the same as that for each component of the $\pi/2$ mode for four segments.

(5) CIRCUIT DESIGN AND LOAD COUPLING

The 2-segment structures are of particular interest for tuning by an external circuit. In order to obtain maximum tuning range the cavities of the 2-segment s.a. structure were shaped to the form of Fig. 6(a).

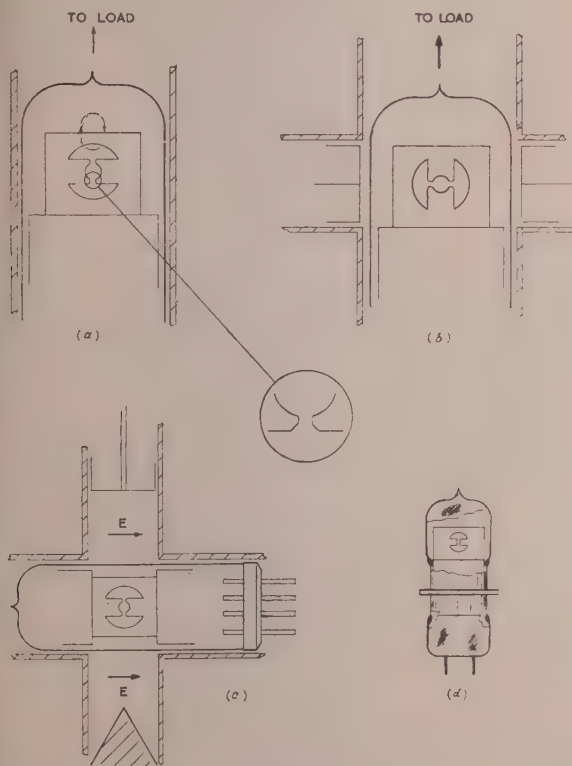


Fig. 6.—Arrangements of 2-segment symmetrical anode.

- (a) Single magnetic coupling.
- (b) Double magnetic coupling.
- (c) For very heavy loading.
- (d) For higher dissipation.

It was also found possible to dispense with the coupling probe of the two types first used, and to couple magnetically from the current circulating round one of the cavities as shown in the Figure. The alternative arrangement at Fig. 6(b) allowed coupling from two resonators to be achieved by rotating the anode through 90° .

(6) PROBLEMS OF VALVE CONSTRUCTION

It will be clear that the achievement of satisfactory performance in valves of the type proposed is greatly dependent on satisfactory mechanical design, and a considerable amount of effort was devoted to problems of valve construction.

(6.1) Cathode Structure Assembly

Experience on other magnetrons had indicated that if a reasonable efficiency and noise performance was to be achieved it would be necessary to use the largest possible ratio between starting current and operating current. With the limited anode dissipation it was therefore essential that the stray current be kept extremely small. That this would depend largely on accurate cathode centring was well known, but whether the necessary accuracy could be achieved without involving high manufacturing skill was uncertain. A number of methods of mounting the cathode were tried with gradually improved results, and in a final arrangement satisfactory performance was obtained. In this the mica mounting pins were moved down to a point near the zero-potential plane for the anode to reduce trouble from possible resonances. They were inserted in the anode block and fixed by deforming the copper around them with a circular tool. Then two metal eyelets were placed on the pins on one side of the anode block. The mica, which now had two clearance holes for the eyelets and a clearance hole for the cathode, was placed in position with a cathode end-shield located by four metal tags loosely mounted on it. This shield, which was made from sheet nickel treated with titanium dioxide, carried the central cathode hole. A jig, fitted with a spigot of the cathode diameter, was then inserted in the anode bore. The eyelets were deformed to clamp to the mica, and the four tags on the end-shield were compressed to fix tightly to the mica. This process was repeated for the second mica, and finally the central jig was removed and the cathode inserted.

A high standard of consistency of cut-off characteristics was achieved by this method, and it was then possible to measure the variation of starting current with anode form (arising from different methods of construction) without cathode eccentricity

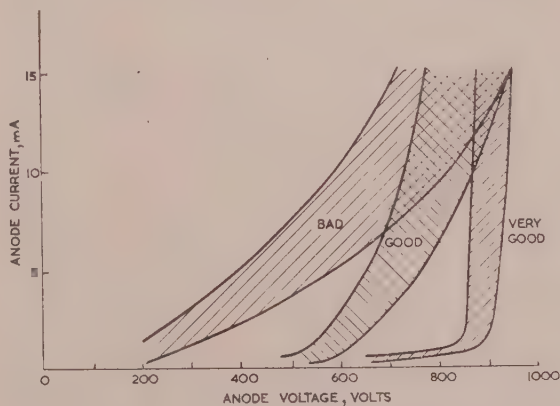


Fig. 7.—Cut-off characteristics for $H = 2400$ oersteds.

masking the effects investigated. Fig. 7 shows the cut-off characteristics obtained with different methods of cathode mounting, the "very good" characteristic being obtained with the method described.

(6.2) Anode Construction

The early anodes of the m.s. type were made by straightforward machining methods, but difficulties were experienced in obtaining the long slots required with sufficient accuracy of form and position. It was found worth while so far as circuit losses were concerned to increase the diameter of the output probe from 0.014 to 0.024 in, the largest wire which could be accommodated in the narrow tongue between the circuit slots.

The first anodes of the s.a. type were made by fabrication, and

the opportunity was taken to form the output coupling probe as part of one of the vanes of which the anode was made, as can be seen in Fig. 1. Later designs used magnetic coupling. Finally, hobbing was used, and this considerably simplified the manufacture, since the anode block could be a straightforward section from a hobbled billet. The need for two machined steps on both sides of the block was also eliminated, and the eyelets holding the mica to the pins were so constructed that the mica was supported away from the block to ensure a high electrical leakage path between anode and cathode.

The anode bore of both 4- and 2-segment blocks was drilled after the anodes were hobbled in early valves, but later it was found possible to construct a hob for forming the anode circuits and central hole together. In the 4-segment case the hob of 80/20 nickel-chromium base alloy was fabricated from two parts; the outer part forming the anode circuits and vanes had a pin of 2 mm diameter inserted centrally to form the anode bore. In the 2-segment anode, owing to the comparatively thin vane forming the segment gaps, it was not possible to use a 2-part fabrication method, and in this case the hob was made by form-grinding a hot-process die steel, care being taken to obtain the minimum radius at the junction of the anode bore and segment tips.

Valves of this type and size (all of 2 mm anode bore and 2.5 mm anode length) are capable of handling an anode dissipation of 8 watts; the temperature of the glass envelope sets the limit to the output of the valve. One method of increasing this dissipation is to provide a high-conductivity connection between the anode and an external cooling element, and this has been tried in a valve having the arrangement indicated in Fig. 6(d). Such a design is likely to increase the cost of the valve somewhat, but the essential simplicity of the electrical design is not affected. The anode dissipation is increased to at least 30 watts, and there is a consequent increase in power output.

(6.3) Cathode

The cathode used in all the experiments described consists of nickel tube of 0.625 mm outside diameter, 0.525 mm internal diameter, spray coated to 0.025 mm thickness over 2 mm length with a mixture of ammonia-precipitated barium and strontium carbonates. The heater supply for 800°C, with the cathode located rigidly to the mica by means of the cathode end-shields, is at 0.3 amp and 6.3 volts. Difficulties due to delamination of the mica around the cathode end-shields were overcome by increasing their thermal emissivity with a coating of titanium dioxide furnace in dry hydrogen to a fine black deposit. The cathode back-bombardment power is normally 3–5% of the anode power; consequently at 8 watts input to the anode a reduction of heater power is necessary. The filament is 0.045 mm molybdenum wire spray-coated with alumina 0.052 mm thick. The eyelets, pins, assembly support pressing and fixings are all of non-magnetic material. A barium getter is mounted near the base and, prior to sealing into the bulb, the base is coated with magnesium oxide. In spite of the close proximity of the getter to the base pins no breakdown occurs at 2000 volts.

(6.4) Processing

The valves are baked on the pump at 380°C for ten minutes to outgas the glass. The complete assembly is then outgassed by eddy-current heating at approximately 750°C. The cathode is flashed at 10 volts to break down the carbonates, after which the getter is flashed and the valve sealed off from the pump.

(7) OPERATION OF 4-SEGMENT M.S. VALVES

Fig. 8 shows a preliminary operating characteristic obtained with a good valve of this type delivering power to a matched load,

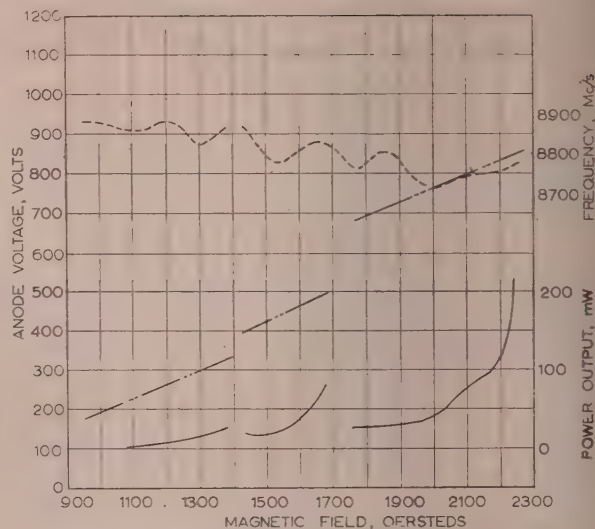


Fig. 8.—Performance of 4-segment m.s. valve.

--- Frequency.
 Voltage.
 — Power.

the coupling being adjusted at each reading to give maximum power output. Power is obtained over a number of discrete ranges of magnetic field corresponding to different component numbers, and there is little variation of frequency as it is essentially that of a single mode of the circuit oscillations. Small changes with operating conditions are due to changes in the admittance of the electron stream.

On correlating the observed ranges of operating voltage with those derived from calculation of threshold voltage² over the range of magnetic field for which a given mode is excited, it appears that operation at high magnetic fields is in the component $n = 5$, and at lower fields in the components $n = 6, 7$ and possibly 8. This is at variance with the analysis already given for this type of structure (Section 3.2.1) from which it is to be expected that either odd or even components only would be excited for a given circuit mode. The appearance of both families may be due to some asymmetry of the structure arising, for example, from inaccurate spacing of the circuit gaps.

(8) OPERATION OF 2-SEGMENT M.S. VALVES

Fig. 9 shows a performance chart for one of these valves when feeding into an output circuit of the type shown in Fig. 15(a). Useful output was obtained but over more limited ranges of operating conditions than for the 4-segment valve. For these tests, coupling of the valve to the load was adjusted for every reading. Comparison of calculated and observed operating voltage showed that the two components observed were $n = 5$ and $n = 6$. No oscillations were observed at lower magnetic field and voltage for modes other than those indicated, and it was concluded that this is simply because the r.f. field in the interaction space is too low for higher-order modes to be excited. As oscillations in both components cease, owing to increase of current to the limit of oscillations, the voltage rises sharply to a value which is rather below that calculated for the critical voltage.

(9) OPERATION OF 4-SEGMENT S.A. VALVES

All valves for which data are given here used magnetic coupling, and Fig. 10 shows a contour diagram of a 4-segment valve coupled to a load through a matching transformer in the way indicated in Fig. 15(a). It will be seen that there are three

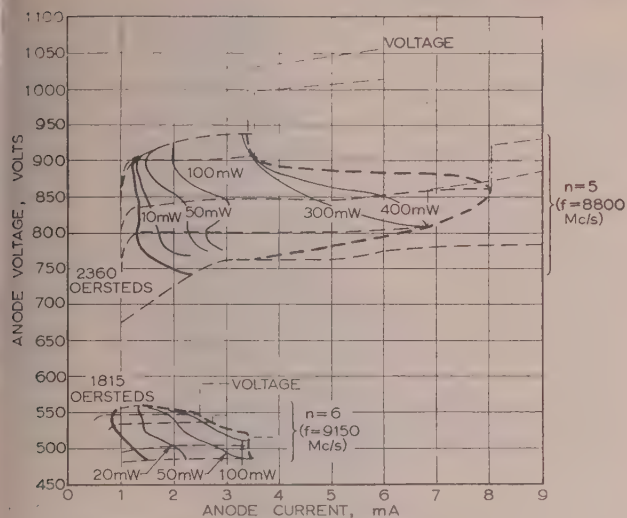


Fig. 9.—Performance of 2-segment m.s. valve.

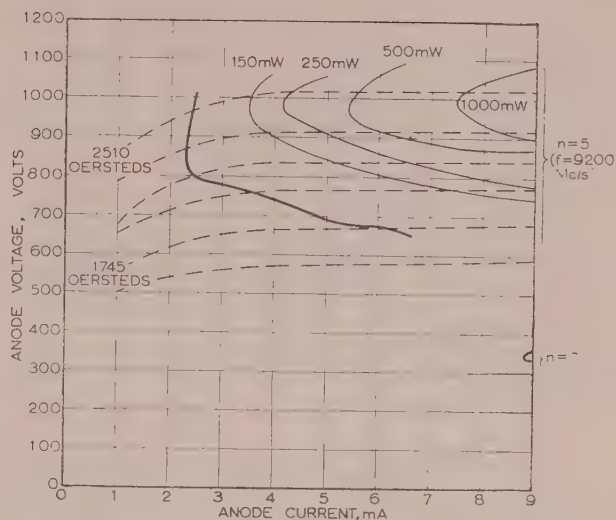


Fig. 11.—Performance of 2-segment s.a. valve.

(10) OPERATION OF 2-SEGMENT S.A. VALVE

Fig. 11 shows a contour diagram under matched load conditions measured in the circuit shown in Fig. 15(a) with operation obtained in the component $n = 5$ only at the normal values of current.

A power of 1 watt is obtained with a magnetic field of about 2450 oersteds, a voltage of 970 volts and a current of 7.5 mA, the efficiency being about 14%. Short-pulse operation to higher powers is also possible.

By increase of current, carried out most conveniently with low-duty-cycle operation, the $n = 7$ component could be detected at low magnetic field, and Fig. 12 shows the frequency/magnetic-

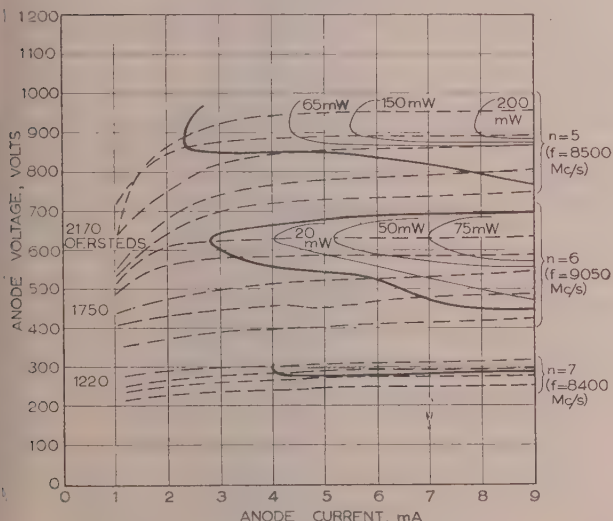


Fig. 10.—Performance of 4-segment s.a. valve.

regions of oscillation, two being of approximately the same frequency. Comparison of observed and calculated operating voltages suggest that these operate through the components $n = 5, 6$ and 7 . The earlier analysis indicates that oscillation of the 4-segment anode in the π mode results in the space components $n = 2, 6, 10$, etc., whilst oscillation in the $\pi/2$ mode results in the space components $n = 1, 3, 5, 7, 9$, etc. Accordingly, it is concluded that the higher-frequency oscillation with $n = 6$ is due to excitation of the resonators in the π mode, the other oscillations of both higher and lower voltage being due to excitation in the $\pi/2$ mode.

It will be noted that the π mode is excited only by one component, others being widely spaced in number.

A c.w. power of 75 mW is generated at a voltage of 600 volts with a current of 9 mA.

Tests have been made to high currents with narrow pulses of low duty cycle. At a magnetic field of 2000 oersteds and voltage of 650 volts a peak power of the order of 1 watt has been obtained with a peak current of 100 mA, with pulses of 2 microsec duration repeated 2000 times a second.

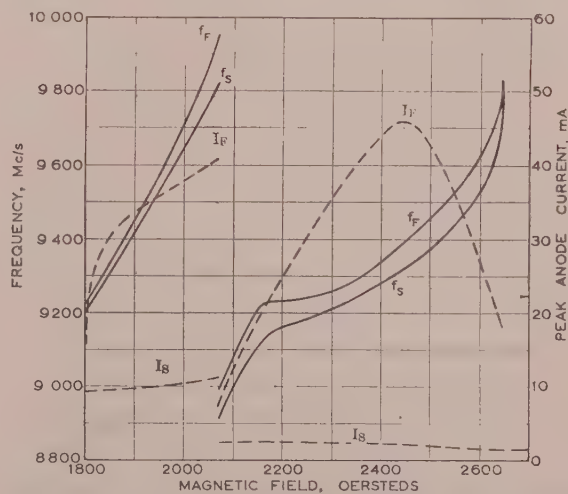


Fig. 12.—Variation of frequency and limiting currents with magnetic field.

field characteristic for both components at currents for the start and finish of oscillations. The variation in frequency over the current range is small compared with that over the magnetic-field range. At fixed current a frequency variation of at least 400 Mc/s may be obtained by variation of magnetic field. It is rather higher than the figure obtained with the 4-segment design.

Valves of this type, for which the above characteristics were

obtained, were constructed with plain anode gaps of width 0.018 in. Earlier valves had gaps of 0.020 in, and these gave consistently lower efficiency (of the order of 3% instead of 14%). Valves made later with the gap still at 0.018 in but the capacitance reduced by chamfering as in Fig. 6(a) have shown higher efficiencies still, and these have been in the neighbourhood of 25% for the component $n = 5$. This change improves the strength of the hob used for anode construction.

Experience in this work of changing gap dimensions and proportions has shown that, as would be expected, it is very necessary to maintain sharp corners at the junction between the anode surface and the interaction gap, otherwise the starting current may be significantly increased and the efficiency consequently decreased.

The change of efficiency resulting from these effects is, however, greater than would be expected from the change in the amplitudes of the r.f. field components.

(11) EFFECT OF LOAD COUPLING ON STARTING AND FINISHING CURRENTS

The influence of circuit loading on the starting and finishing currents (I_S and I_F) is shown generally in Fig. 13(a) for the valves

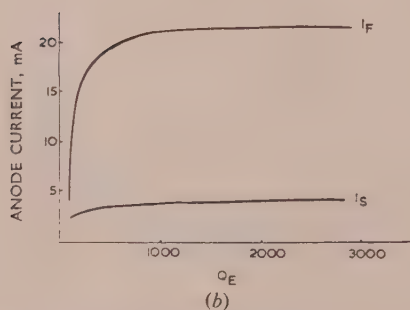
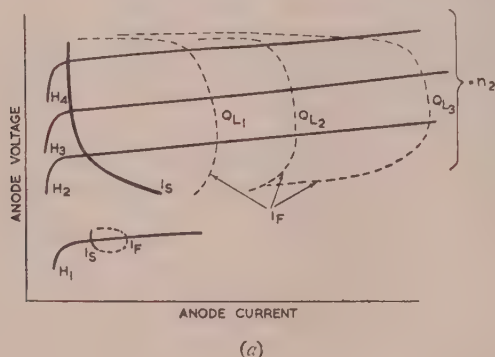


Fig. 13.—Influence of load coupling on I_S and I_F .

(a) Contour diagram.
(b) Variation of I_S and I_F with Q_E .

whose characteristics have been described. These valves have a ratio r_c/r_a of cathode to anode diameter of 0.31, and it is possible that variation of this ratio will modify these characteristics.

In the higher field range (H_2, H_3, H_4) where operation is in the component n_2 , the starting current falls somewhat with increasing magnetic field at a given overall circuit Q-factor of Q_L . It is little affected by circuit loading, and this behaviour is consistent with the presence of significant (though small) stray currents whose magnitude is not greatly dependent on the presence of normal oscillations.

On the other hand, the current at which oscillations cease, I_F , varies significantly and is strongly dependent on Q_L . It appears that I_F is determined entirely by the magnitude of the r.f. field coupling to the electron wave. At low magnetic field this is small, because of low efficiency, and it increases markedly as efficiency is increased with magnetic field. At high magnetic field, however, owing to shrinkage of the electron cloud the r.f. coupling field falls again, even though the r.f. field may be large at the surface of the anode.

This behaviour is repeated for the higher-order component n_1 at the lower magnetic field H_1 , although over a much smaller range of parameters.

Fig. 13(b) shows values of both currents I_S and I_F at a fixed magnetic field plotted as a function of Q_E , this being the component of Q_L which is due to the external load only. These data were taken on a 2-segment s.a. valve having a "side" coupling to tuning waveguides (see Section 14). It will be noted that I_F increases markedly with Q_E but reaches a limiting maximum when Q_E becomes so large that it has little effect on the overall loss factor Q_L .

(12) FREQUENCY CHANGE WITH CURRENT

It has been found that the change of frequency with current (frequency pushing) is much greater with spatial-harmonic operation than with fundamental operation. Measurements were made of the frequency change with current of a 2-segment s.a. valve operating at a magnetic field of 2300 oersteds in the $n = 5$ component when modulated over the current range of 1–9.5 mA. The valve was coupled to the output waveguide sufficiently tightly to give maximum power output. It was found that the rate of change of frequency with current varied little over the range and was of the order of 14.5 Mc/s/mA. The output waveguide was then mismatched by a probe giving a standing-wave ratio of up to 1.75, and the rate of change of frequency with current was measured as the mismatching probe was moved along through a guide length of greater than a half wavelength for different values of the mismatch. It was, however, necessary to limit the current to 6.5 mA for the mismatch setting giving the highest load conductance at the valve.

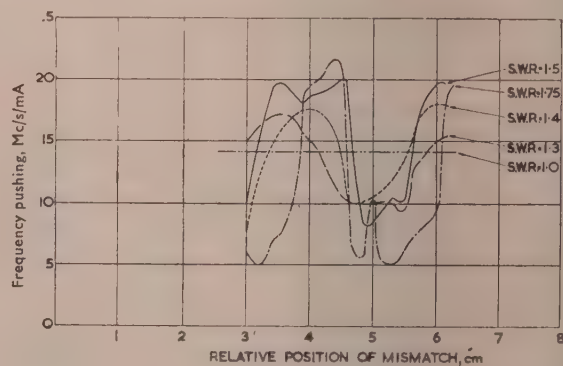


Fig. 14.—Variation of frequency pushing with circuit loading.

Fig. 14 shows the values obtained over the range of loading parameters, the high figures corresponding to maximum load conductance and the low figures to low conductance. A peak value of 21.5 Mc/s/mA is obtained with a peak frequency excursion of the order of 170 Mc/s (at slightly less than maximum conductance). For a power variation of not more than 3 dB from the peak value a frequency excursion of rather less than half the above value is obtained.

(13) FREQUENCY CHANGE WITH CURRENT UNDER CONDITIONS OF EXTREME CIRCUIT LOADING

One possibility offered by the simple resonant system used in this work is that of investigation of magnetron performance in the condition where coupling of the load to the resonant circuits is so tight that the impedance measured at the interaction gaps loses its resonant properties and becomes resistive. Under this condition general considerations suggest, and work elsewhere has confirmed, that it is possible to vary the frequency generated over a wide range by alteration of the velocity of the electron cloud circling the cathode, by variation of applied voltage, magnetic field, or both. Measurements by Wilbur and Peters³ at frequencies of the order of 550 Mc/s showed that variation of frequency over a range of the order of 2:1 could be obtained with power output of upwards of 20 watts. In order to obtain a signal with a line spectrum, reduction of emission to either partial or full temperature limitation was required.

Similar performance was obtained by Guénard and Huber,⁴ but in their case no requirement was found for emission limitation.

Only preliminary measurements have been made on valves of the 2-segment s.a. type with the structure modified as indicated in Fig. 6(c) to increase the coupling to the load by adding an additional choke. With this arrangement the normal oscillations of substantially fixed frequency could not be obtained, but on operating the valve with a varying anode voltage at 50 c/s it was observed that output was obtained over a range of frequency.

Preliminary measurements were made on a number of valves of this type and in all cases a frequency change of at least 1 000 Mc/s was obtained with variation of anode current, the mean frequency being of the order of 9 500 Mc/s. By altering the arrangement for coupling the valve to the load circuit so that the low-frequency cut-off arising from the output waveguide was eliminated, it was found possible to detect output energy over the range 10 000–5 000 Mc/s by exploring over an extreme range of current. The power measured in this case was significantly less than that with normal operation; about one milliwatt was obtained with a peak current of 40 mA.

In these experiments no sign was seen of the requirement for a particular emission condition. A clean spectrum was obtained with normal space-charge limitation of emission, and this is of particular importance for the practical utilization of this type of operation. On the other hand, the absence of temperature limitation results in large variation in anode current and power over a useful frequency range.

(14) MECHANICAL TUNING

One of the advantages envisaged from the start of this work, for the structure described in this paper, was the possibility of wide-range mechanical tuning by means of an externally coupled reactance. The single mode of operation of the 2-segment valve, in particular, should allow considerable tuning ranges to be achieved.

A number of different tuning circuits and arrangements have been tried and are indicated schematically in Fig. 15.

An arrangement used for investigation of the tuning performance of the 2-segment m.s. design is shown in Fig. 15(b). No deliberate damping of the tuning circuit by a load was employed, and the valve projects into the resonant circuit formed by a piece of rectangular waveguide which is terminated by a tuning piston. A wavemeter coupled loosely to the cathode leads of the valve showed that a tuning range of over 1 000 Mc/s could be obtained in the higher-voltage mode. In this case, although no power was delivered to a useful load, the power transmitted through the wavemeter from the heater leads was of the order of 1 mW, so that in fact a power of at least 10 mW was probably being radiated.

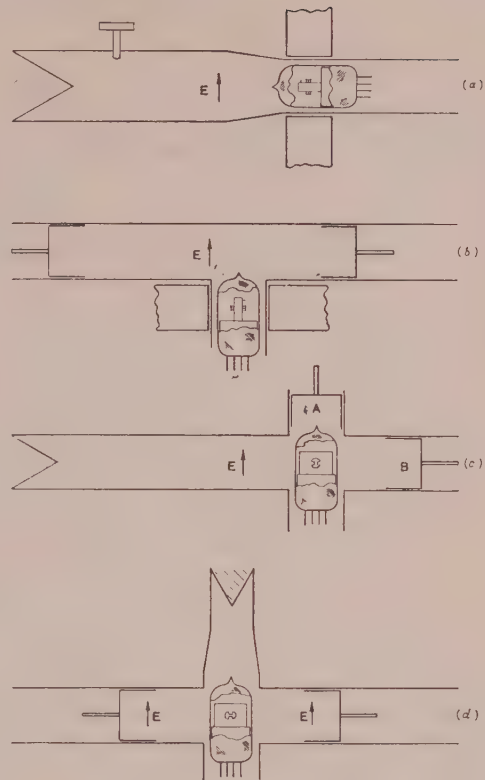


Fig. 15.—Coupling and tuning circuits.

The arrangement which has been found most convenient in use with 2-segment s.a. valves having end coupling is that shown in Fig. 15(c). In this the valve is coupled by its magnetic field to the tuning waveguide, A, the coupling being adjusted by the degree of insertion of the valve into the circuit and by the position of piston B. The load circuit is coupled by series connection to the output waveguide.

Fig. 16 shows characteristics of frequency and power measured as a function of position of the tuning piston with piston B set to give either maximum power output at some part of the range or minimum power variation over the range. In the first case a tuning range of $7\frac{1}{2}\%$ of the mean frequency is obtained with a minimum power of 170 mW and maximum power of 660 mW. With the piston set to give minimum power variation, a tuning range of 4.3% is obtained with a minimum power of 200 mW and a maximum of 235 mW. These figures apply to operation with constant magnetic field. It will be expected from the characteristics shown in Fig. 12 that some further increase of tuning range of the order 200–300 Mc/s may be achieved by variation of magnetic field.

The arrangement used in the loading experiments described in Section 12 is shown in Fig. 15(d). Here a 2-segment valve with double coupling, as shown in Fig. 6(b), is used. By movement of the tuning plungers symmetrically relative to axes of the valve, tuning may be obtained with zero coupling to the load. By displacing one tuning piston by a given amount from its setting for a particular frequency, and the other piston by a similar but negative amount, the frequency may be left substantially unchanged, although the coupling to the load may be increased to a desired value. Thus, independent control of frequency and load coupling is achieved, and a wide range of loading may be covered.

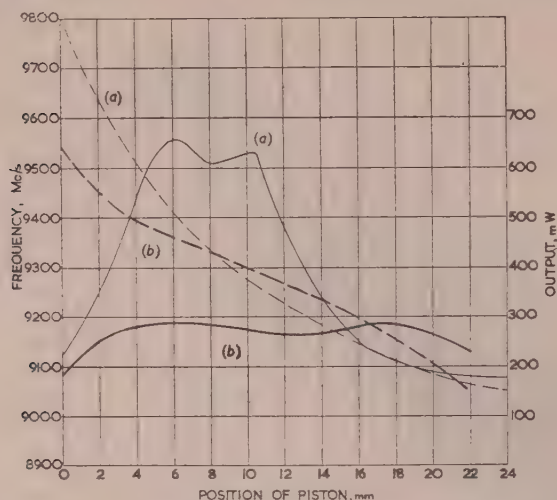


Fig. 16.—Tuning characteristics of 2-segment s.a. valve:

- (a) Adjusted for maximum power output.
 (b) Adjusted for minimum power variation.
 ——— Power.
 - - - - - Frequency.

Experience obtained in mechanical tuning of valves of this type is so far only preliminary, and it is considered that wider tuning ranges may be eventually obtained with sufficient attention to the design of tuning circuits. More rapid control of frequency may be obtained by the use of a phase shifter of the ferrite type, with which the change of electrical length of the tuning waveguide may be effected by change of the magnitude of polarizing magnetic field.

(15) NOISE PERFORMANCE

Few data have so far been obtained on the noise performance, but preliminary measurements of amplitude-modulation noise indicate that the level is low enough for the valve to be used as a local oscillator in a superheterodyne receiver with a bandwidth of a few megacycles per second without degradation of receiver sensitivity due to local-oscillator noise.

(16) OSCILLATION BUILD-UP TIME

One parameter which is of interest in low-power microwave generators is that of time of oscillation build-up. In some applications it is desirable for this to be as short as possible. The use of an anode structure of the type here described appears to allow the possibility of very short build-up time with minimum

time jitter, since interfering effects due to the generation of unwanted modes are minimized and, in fact, are not present in the 2-segment structure. No detailed measurements have been carried out, but preliminary indications are that very rapid and stable starting is achieved when pulses considerably less than 0.1 microsec long are generated, the limit of build-up time being so far set by the shape of the current pulse which can be generated.

(17) LIFE

The most probable cause of failure in valves of this type is loss of cathode emission. With the limit of anode dissipation imposed by the structure (see Section 8) the maximum emission required is 0.2 amp/cm at a frequency around 10 000 Mc/s. In practice a value somewhat lower than this will normally be used. Providing, therefore, that the cathode temperature does not exceed 1000° K, thorough outgassing and adequate gettering should ensure a long life.

(18) CONCLUSION

It appears that the use of spatial-harmonic operation can lead to designs of magnetron having attractive characteristics of performance, design and construction.

Work is still being carried out, but the results reported here, although preliminary in nature, are sufficient to indicate the potentialities of this type of operation.

(19) ACKNOWLEDGMENTS

The authors wish to refer to the contribution of Mr. E. B. Callick, of the Admiralty Department of Physical Research, whose assistance in planning the mechanical design of the first valves is particularly acknowledged.

The authors are also indebted to the Admiralty for permission to publish the paper.

(20) REFERENCES

- (1) WILLSHAW, W. E., and RUSHFORTH, L.: "The High Power Pulsed Magnetron: An Outline of Operation," *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 180.
- (2) COLLINS, G. B.: "Microwave Magnetrons," Chapter 1 (McGraw-Hill, 1948).
- (3) DONAL, J. S.: "Modulation of Continuous-Wave Magnetrons," *Advances in Electronics IV* (Academic Press, New York, 1952).
- (4) GUÉNARD, P., and HUBER, H.: "Étude expérimentale de l'interaction par ondes de charge d'espace au sein d'un faisceau électronique se déplaçant dans des champs électrique et magnétique croisés," *Annales de Radio-électricité*, 1952, **7**, p. 252.

THE DETERMINATION OF THE TRUE SIDE-LOBE LEVEL OF LONG BROADSIDE ARRAYS FROM RADIATION-PATTERN MEASUREMENTS MADE IN THE FRESNEL REGION

By R. H. T. BATES, B.Sc.(Eng.), Graduate, and J. ELLIOTT, M.Sc.

(The paper was first received 18th October, and in revised form 16th December, 1955. It was published as an INSTITUTION MONOGRAPH in March, 1956.)

SUMMARY

The radiation pattern in the Fresnel region is expressed as a function of the far-field polar diagram. Following this it is shown how the true side-lobe level may be inferred from amplitude radiation-pattern measurements made in the Fresnel region. This results in considerable shortening of the range at which polar-diagram measurements may be made.

LIST OF PRINCIPAL SYMBOLS

- 2a = Interval in ψ between zeros of main beam of polar diagram.
- $C(z) = \int_0^z \cos \frac{\pi t^2}{2} dt$ = Fresnel's cosine integral.
- $S(z) = \int_0^z \sin \frac{\pi t^2}{2} dt$ = Fresnel's sine integral.
- D = Distance from measuring point to centre of array.
- $E(\theta) = B(\psi)$ = Radiation pattern in Fresnel region.
- $P(\theta) = A(\psi)$ = Polar diagram.
- $H(z)$ = Function defined in text.
- $I(x)$ = Amplitude distribution of sources along array.
- J_n = Integral defined in Appendix.
- $k = 2\pi/\lambda$ = Wave number of power radiated or received by array.
- L_{mn} = Level of n th side lobe of m th polar diagram.
- l = Half-length of array.
- O = Centre point of array.
- n = An integer greater than zero.
- Q = Measuring point.
- r = Distance from measuring point to any point on array.
- x = Array co-ordinate.
- α, μ, ν = Parameters.
- β_n, γ_n = Constants defined in Appendix.
- λ = Free-space wavelength of power radiated or received by array.
- θ = Angle to array measured from normal to array.
- 2ϕ = Angular width between zeros of main beam of polar diagram.
- $\psi = k \sin \theta$.
- ζ, z = Dummy variables.

(1) INTRODUCTION

The electromagnetic field associated with any aerial system can be roughly divided into three regions, namely

- (a) Induction or near field.
- (b) Fresnel region.
- (c) Fraunhofer region or far field.

A radiation pattern is defined as the relative field strength measured at a fixed point as the aerial is rotated, and the shape of this pattern is independent of distance from the aerial only in region (c). There is no definite dividing line between regions (b) and (c), but there are reasonable criteria for determining whether the measuring point is in the far field. One of these,¹ which applies to aeri-als whose sources are situated in a straight line, is that the distance from the measuring point to any part of the array should be greater than $8l^2/\lambda$.

For many practicable arrays the measuring point would have to be several miles distant. This would not be an insurmountable difficulty, but it could prove inconvenient; e.g. a suitable site for the measurement would not be easy to find.

To determine completely the polar diagram from a radiation-pattern measurement taken in the Fresnel region it is necessary to discover the relative phase of the field as well as its amplitude. However, it will be shown in the paper how the main characteristics of the polar diagram may be inferred from the Fresnel-region amplitude-radiation pattern alone. The measurement of relative amplitude can be done simply and quickly, whereas the measurement of relative phase is altogether more tedious.

The results given in the paper apply to linear arrays whose main beams are within about 30° of the broadside direction.

(2) THEORETICAL INTRODUCTION

Only arrays whose sources are situated in a straight line are considered in the paper, and the sources are considered to be similarly polarized.

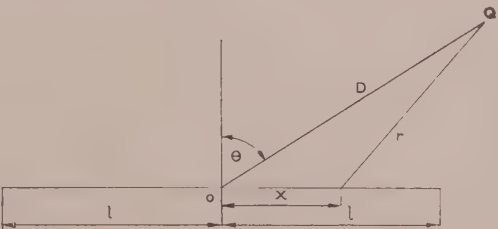


Fig. 1.—Geometry of array and measuring point.

Fig. 1 describes the geometry relating the array, which exists between $x = -l$ and $x = l$, and the measuring point Q. The radiation diagram, $E(\theta)$, is recorded at Q while the array is rotated about O in the plane of the paper.

From the potential theory² of the electromagnetic field the radiation diagram of the field strength polarized parallel to the sources is given by

$$E(\theta) = \int_{-l}^l I(x) \frac{e^{-jkr}}{r} dx \dots\dots\dots (1)$$

Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Bates is with Decca Radar, Ltd., and Mr. Elliott is with Vickers-Armstrongs (Aircraft), Ltd.

This formula is exact, provided that the amplitude distribution consists of only electric or magnetic sources. If the distribution is of a mixed kind, the two types of source must be treated separately and the results combined at the end of the analysis.

Since we are interested only in relative values there is no need to insert any constant of proportionality in eqn. (1).

It is necessary to determine how large D must be for Q to be in the Fresnel region. There are two conditions, namely

- (a) Variations of r with x in the denominator of the integrand on the right-hand side of eqn. (1) may be neglected.
- (b) r in the exponential may be expressed as

$$r \simeq D + \frac{x^2}{2D} - x \sin \theta; \text{ for } |\theta| < \text{about } 30^\circ$$

This expression is obtained by expanding r in a binomial series.

Applying these conditions to eqn. (1) gives

$$E(\theta) = \frac{e^{-jkD}}{D} \int_{-\infty}^{\infty} I(x) e^{-jkx^2/2D} e^{jkx \sin \theta} dx$$

The limits have been changed to $\pm \infty$ for analytical convenience. Since $I(x)$ exists only in the interval $-l < x < l$, this equation is the same as eqn. (1).

Writing

$$\psi = k \sin \theta \quad . \quad . \quad . \quad (2)$$

and discarding constants of proportionality gives

$$B(\psi) = E(\theta) = \int_{-\infty}^{\infty} I(x) e^{-jkx^2/2D} e^{j\psi x} dx \quad . \quad . \quad (3)$$

In the far field $kx^2/2D \ll \psi x$, and so the polar diagram can be written as

$$A(\psi) = P(\theta) = \int_{-\infty}^{\infty} I(x) e^{j\psi x} dx \quad . \quad . \quad (4)$$

Eqn. (4) is true² for all values of θ .

We wish to show how the main characteristics of a polar diagram may be inferred from measurements taken in the Fresnel region. It is unlikely that a general relation could be derived from eqn. (3). The aperture distribution, $I(x)$, can be infinitely varied in practice, and approximating it by the first few terms of an orthonormal series appears to give results which are too complicated to be combined into a universal formula.

There is one thing in which the majority of amplitude distributions applied to long arrays are similar: they produce polar diagrams whose shapes are much the same. This applies adequately to polar diagrams whose main beams are within some 30° of the normal of the array. It is also usual for the first side-lobe to be so placed that, if $2a$ is the interval in ψ between the zeros of the main beam, $3a/2$ is approximately the interval in ψ between the main beam maximum and the first side-lobe maximum. The most apparent difference between polar diagrams is the first side-lobe level, which is a parameter which can be varied easily.

Following these arguments, $B(\psi)$ should be expressed in such a way that it is a function of $A(\psi)$ rather than $I(x)$. The clue to how to do this is given by eqn. (4), where it will be seen that $A(\psi)$ and $I(x)$ are a Fourier-transform pair. The analysis is given in the next Section.

(3) RADIATION PATTERN EXPRESSED AS A FUNCTION OF POLAR DIAGRAM

Proceeding formally from eqn. (4),

$$I(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\zeta) e^{-j\zeta x} d\zeta \quad . \quad . \quad (5)$$

Substituting eqn. (5) in eqn. (3),

$$B(\psi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\zeta) e^{j(\psi - \zeta)x} e^{-jkx^2/2D} d\zeta dx \quad . \quad (6)$$

Consider

$$H(z) = \int_{-\infty}^{\infty} e^{-jkx^2/2D} e^{jzx} dx \quad . \quad . \quad (7)$$

This is another Fourier transform. From Reference 3 we see that

$$H(z) = \frac{(1-j)\pi^{1/2} D^{1/2}}{k^{1/2}} e^{jDz^2/2k} \quad . \quad . \quad (8)$$

$$B(\psi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} A(\zeta) H(\psi - \zeta) d\zeta \quad . \quad . \quad (9)$$

Substituting eqn. (8) in eqn. (9),

$$B(\psi) = \frac{(1-j)D^{1/2}}{2\pi^{1/2}k^{1/2}} \int_{-\infty}^{\infty} A(\zeta) e^{jD(\psi - \zeta)^2/2k} d\zeta \quad . \quad (10)$$

The radiation pattern in the Fresnel region has now been expressed as a function of the polar diagram.

There is an interesting and simple check on the work leading to eqn. (10). When D is large, i.e. the measuring point is in the far field, the principle of stationary phase⁴ may be applied to the integral in eqn. (10).

It will be found that

$$\lim_{D \rightarrow \infty} B(\psi) = A(\psi)$$

(4) CHANGE IN SIDE-LOBE LEVEL WITH RANGE

The purpose of most long broadside arrays is to produce a polar diagram with a single narrow main beam. The two most important features of such polar diagrams are the width of the main beam, either between 3 dB or zero points, and the level of the largest side-lobe, which is often the first, i.e. that nearest the main beam.

The following assumptions about the form of the polar diagram will be made so that a general, though necessarily approximate, theory of the variation of side-lobe level with range can be constructed:

(a) Width of main beams between zeros = $2a$, measured in units of ψ .

(b) Interval in ψ between main beam maximum and first side lobe maximum = $3a/2$.

(c) Interval in ψ between adjacent side lobes = a .

As was suggested in Section 3, departures from these assumptions will be slight in practice.

There is a certain property of the integral in eqn. (10) which is useful for establishing the general result which we seek. Consider $B(\psi)$ for the two polar diagrams $A(\psi)$ and $A(\psi - \alpha)$. From eqn. (10),

$$B_I(\psi) = \frac{(1-j)D^{1/2}}{2\pi^{1/2}k^{1/2}} \int_{-\infty}^{\infty} A(\zeta) e^{jD(\psi - \zeta)^2/2k} d\zeta$$

$$B_{II}(\psi) = \frac{(1-j)D^{1/2}}{2\pi^{1/2}k^{1/2}} \int_{-\infty}^{\infty} A(\zeta - \alpha) e^{jD(\zeta - \psi)^2/2k} d\zeta$$

$$= \frac{(1-j)D^{1/2}}{2\pi^{1/2}k^{1/2}} \int_{-\infty}^{\infty} A(z) e^{jD(z - \psi + \alpha)^2/2k} dz$$

So

$$B_I(\psi) = B_{II}(\psi - \alpha) \quad . \quad . \quad (11)$$

Thus, if two polar diagrams are similar but displaced in ψ by an amount α , their radiation patterns in the Fresnel region are also similar and displaced in ψ by the same amount α . This means that any results derived for polar diagrams whose main beams are centred on $\psi = 0$ can be applied to similar polar diagrams displaced in ψ by an amount equivalent to $|\theta|$ less than about 30° . This restriction on the values of θ and ψ applies throughout the paper. The word "similar" has been used here in the same sense as it is used in Euclidean geometry.

Formulae approximating polar diagrams can be inserted into the integral in eqn. (10) and the radiation patterns derived. In order to make the results as general as possible, and at the same time to keep the computations within reasonable bounds, we will calculate $B(\psi)$ for two types of polar diagram, designated $A_1(\psi)$ and $A_2(\psi)$:

$$\left. \begin{aligned} A_1(\psi) &= \frac{\sin \frac{\pi\psi}{a}}{\frac{\pi\psi}{a}}; & -\infty < \psi < -a \\ &= \mu \cos \frac{\pi\psi}{2a} + \frac{\sin \frac{\pi\psi}{a}}{\frac{\pi\psi}{a}}; & -a < \psi < a \\ &= \frac{\sin \frac{\pi\psi}{a}}{\frac{\pi\psi}{a}}; & a < \psi < \infty \end{aligned} \right\} \quad (12)$$

$$\left. \begin{aligned} A_2(\psi) &= -\nu \sin \frac{\pi\psi}{a}; & -\infty < \psi < -a \\ &= \cos \frac{\pi\psi}{2a}; & -a < \psi < a \\ &= \nu \sin \frac{\pi\psi}{a}; & a < \psi < \infty \end{aligned} \right\} \quad (13)$$

These polar diagrams can be classified roughly as the two extreme members of a set. The side lobes of $A_1(\psi)$ decrease in height as they recede from the main beam at a rate which will be seldom exceeded in practice. When $\mu = 0$, $A_1(\psi)$ is the polar diagram due to a uniform amplitude distribution. The side lobes of $A_2(\psi)$ are all equal in height. This is a property of polar diagrams produced by Dolph-Chebyshev⁵ amplitude distributions which optimize the relationship between beam width and side-lobe level; this level is varied by altering the parameters μ and ν .

Substituting eqns. (12) and (13) in eqn. (10) yields

$$B_1(\psi) = J_1 + \mu J_3 \quad (14)$$

$$B_2(\psi) = J_3 + \nu[J_5 + J_6] \quad (15)$$

The functions J_n are defined in the Appendix, where the necessary integrations are described.

Examination of the expressions for $A_1(\psi)$ and $A_2(\psi)$ shows that the side-lobe maxima occur at $\psi = \frac{1}{2}(2n+1)a$, n being an integer greater than zero. The level of the n th side lobe relative to the main beam, measured in the Fresnel region for the two specified polar diagrams, will be

$$L_{1n} = \left| \frac{B_1\left(\frac{2n+1}{2}a\right)}{B_1(0)} \right| \quad (16)$$

$$L_{2n} = \left| \frac{B_2\left(\frac{2n+1}{2}a\right)}{B_2(0)} \right| \quad (17)$$

Comparison of eqns. (16) and (17) with (14) and (15) shows that L_{1n} and L_{2n} can be expressed as functions of $a^2 D/k$ alone. Reference to eqn. (1) shows that $a = k \sin \phi$; where $2a$ and 2ϕ are the width of the main beam between its zeros measured in units of ψ and θ respectively. Since $k = 2\pi/\lambda$, $a^2 D/k = (2\pi D/\lambda) \sin^2 \phi$, which is a dimensionless quantity.

The most important side-lobe is usually the first. This being so, L_{11} and L_{21} have been calculated for various values of the true side-lobe level. The values of the Fresnel integrals were found in Reference 6, and Fig. 2 shows the results of the cal-

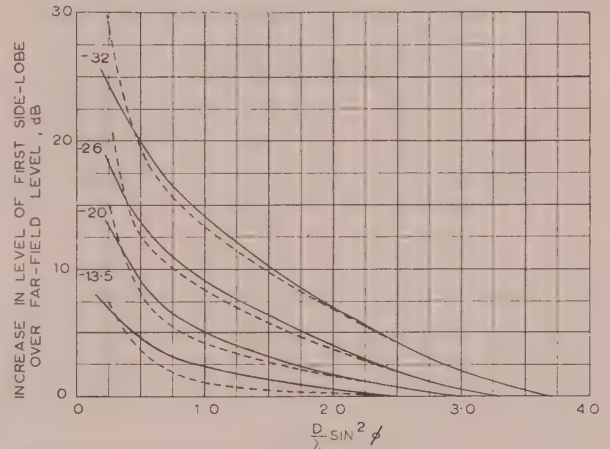


Fig. 2.—Comparison of first side-lobe levels in the Fresnel region for two polar diagrams.

--- $A_1(\psi)$
--- $A_2(\psi)$
Parameter: True level of first side lobe, dB.

culations. The graphs for L_{11} and L_{21} are nearly the same, except at small values of $(D/\lambda) \sin^2 \phi$. Since $A_1(\psi)$ and $A_2(\psi)$ can roughly be considered as the extreme members of the set of polar diagrams peculiar to long broadside arrays, the differences between the graphs for L_{11} and L_{21} could be taken as the possible error if the mean results in Fig. 2 were applied to some polar diagram other than $A_1(\psi)$ or $A_2(\psi)$.

While calculating L_{1n} and L_{2n} it has been assumed that the side-lobe maxima stay in the same positions while D/λ is varied. This is not exact, but side-lobe maxima are generally broad. There will be little error if the calculations are made at values of ψ near to those at which the maxima occur. It is shown in the next Section that the beam width varies slowly with D/λ . Thus the side-lobe maxima will probably change their positions very little for the values of $(D/\lambda) \sin^2 \phi$ used in Fig. 2.

(5) PRACTICAL USE OF THE THEORY

The variation of side-lobe level with range is expressed solely in terms of D , λ and the width of the main beam of the polar diagram; the physical length of the array is not used in the argument. This means that the width of the main beam between its zeros must be determined before the theory can be used in practice. The following hypothesis is necessary.

Consider those points of the amplitude polar diagram at which there are either minima or zeros. In the Fresnel-region amplitude radiation pattern the zeros will become minima and these will be at a higher level with respect to the main-beam

maximum than they were in the polar diagram. However, the positions of the minima will not be changed appreciably.

It is suggested that the above will hold in general for the types of polar diagram considered in the paper. To provide some sort of verification for the hypothesis, the amplitude-radiation patterns in the neighbourhood of the first zero of a particular polar diagram, for several values of $(D/\lambda) \sin^2 \phi$, are shown in Fig. 3. The curve marked $(D/\lambda) \sin^2 \phi = \infty$ is the

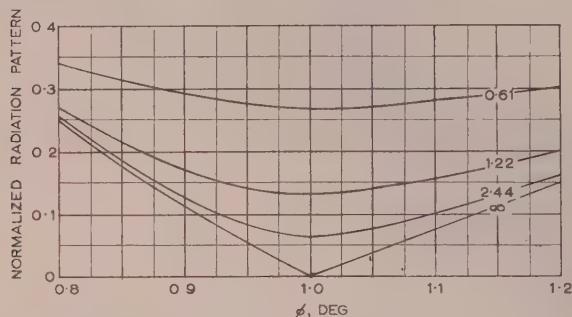


Fig. 3.—Amplitude radiation patterns in neighbourhood of first

$$\text{minimum, for } A(\psi) = \frac{\sin \pi \psi}{a} \bigg/ \frac{\pi \psi}{a}$$

Parameter: $(D/\lambda) \sin^2 \phi$.

polar diagram. It can be seen that the positions of the minima are practically identical, although their levels change appreciably. The patterns are normalized so that the value of the amplitude of the main beam is unity. The results given in Fig. 3 cannot be classed as a general theorem, but it seems probable that the hypothesis will hold in general.

The following statement can now be made: The width of the main beam of the polar diagram between its zeros or minima is the same as the width of the main beam, measured in the Fresnel region, between its minima.

Fig. 4 has been constructed for practical use. The procedure for finding the true side-lobe level is:

- Measure the wavelength, λ , of the radiation used.
- Measure the distance, D , between the measuring point and the centre of the array.
- Measure an amplitude radiation pattern.
- Note the width of the main beam, 2ϕ , between minima.
- Calculate $(D/\lambda) \sin^2 \phi$.
- Read off the true side-lobe level from Fig. 4, having noted the measured level.

Fig. 5 gives the possible error inherent in using Fig. 4. This error has been derived from Fig. 2 on the assumption that $A_1(\psi)$ and $A_2(\psi)$ are the extreme members of the set of polar diagrams peculiar to long broadside arrays. From Fig. 2 it can be seen that the curves for $A_1(\psi)$ and $A_2(\psi)$ cross in the neighbourhood of $(D/\lambda) \sin^2 \phi = 0.4$. It is thought that the curves for other possible polar diagrams need not all meet in this neighbourhood. Therefore, the error curves have been assumed to increase steadily as $(D/\lambda) \sin^2 \phi$ tends to zero.

By examining Figs. 4 and 5 it can be seen that the theory is of little use for values of $(D/\lambda) \sin^2 \phi$ less than about 0.5. There is a physical reason for this. For $(D/\lambda) \sin^2 \phi$ greater than approximately 0.5 the array will consist of a single Fresnel zone. The meaning of this last statement is as follows: Referring to Fig. 1, when $\theta = 0$ the aperture consists of a single Fresnel zone if the greatest value of r does not exceed D by more than $\lambda/2$. If D/λ is so small that the aperture consists of more than one Fresnel zone, the radiation pattern may bear very little resemblance to the polar diagram, which means that the theory of

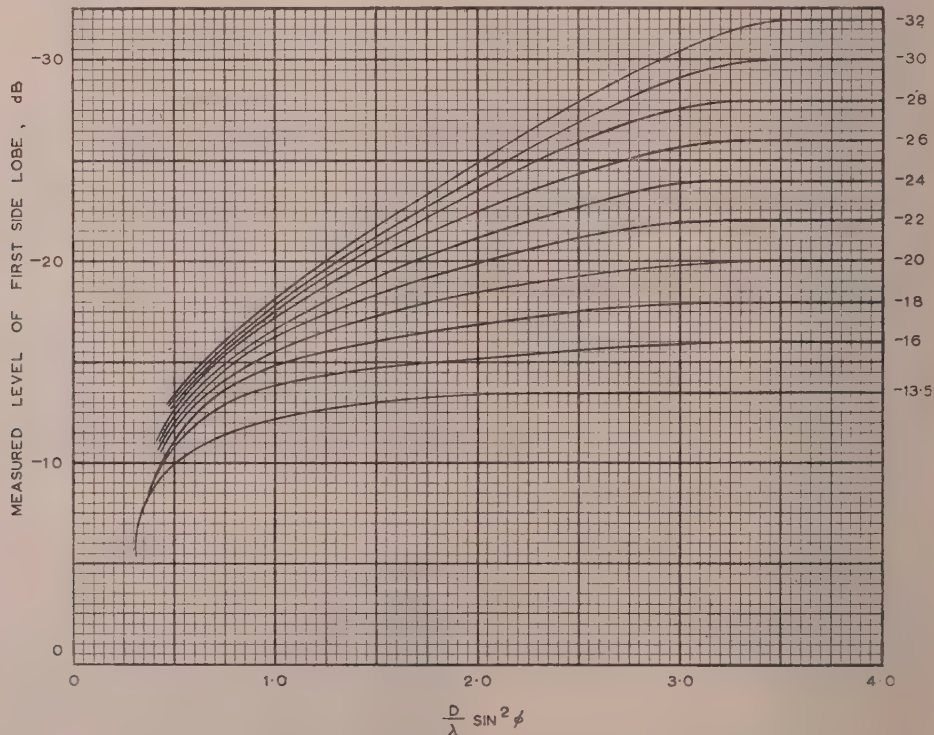


Fig. 4.—Variation of first side-lobe level with range.

Parameter: True level of first side lobe, dB.

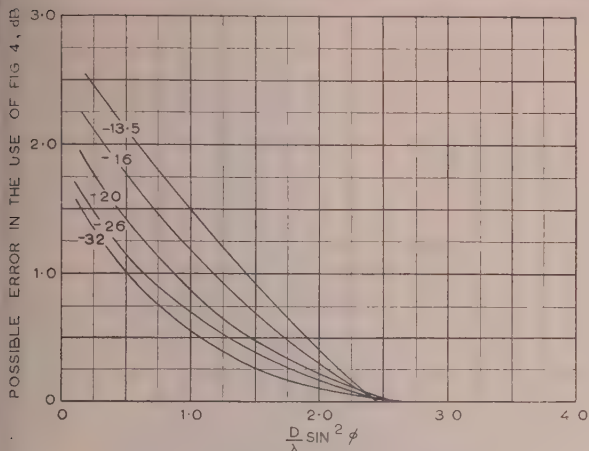


Fig. 5.—Possible error in the use of Fig. 4.
Parameter: True level of first side lobe, dB.

the paper cannot be applied. A good account of Fresnel zones will be found in Reference 7.

There is a type of polar diagram occurring quite frequently in practice to which the results of the paper might be thought not to apply. It may happen that the main beam and the first several side-lobes of a polar diagram fall within the set bounded by $A_1(\psi)$ and $A_2(\psi)$, but the side lobes in the end-fire directions may be comparatively large. This will occur if the aperture distribution consists of discrete sources spaced nearly a wavelength apart. However, it is suggested that, so long as only the last two or three side lobes are larger than the first, the results in Fig. 4 will still apply. Consider eqn. (10). The function $e^{jD(\zeta - \psi)^2/2k}$ varies slowly about $\zeta = \psi$. If D/k is comparatively large the function will oscillate rapidly for values of ζ well away from $\zeta = \psi$. The results shown in Fig. 4 are derived for values of ψ in the neighbourhood of zero. The contribution to the integral in eqn. (10) from end-fire side lobes will then be small. It is thought that this reasoning will apply adequately when the array consists of a single Fresnel zone. This is the condition for which the results of the paper apply.

(6) CONCLUSIONS

An approximate theory of the variation of side-lobe level with range has been developed for long broadside arrays. The results apply when the main beams are centred at not more than 30° from the normal to the array and when the array consists of a single Fresnel zone. A universal chart, Fig. 4, has been constructed to show the true level of the first side lobe when the range of the measuring point and the measured side-lobe level are known. Fig. 5 gives the possible error inherent in Fig. 4.

The purpose of the paper has been to show that the main characteristics of polar diagrams of long broadside arrays may be inferred from amplitude radiation patterns taken in the Fresnel region. The advantage of being able to do this is that the range at which measurements have to be made is considerably reduced.

(7) ACKNOWLEDGMENT

The permission of Vickers-Armstrongs (Aircraft), Ltd. to publish this paper is gratefully acknowledged.

(8) REFERENCES

- (1) KRAUSS, J. D.: "Antennas" (McGraw-Hill, New York, 1950), p. 117.
- (2) STRATTON, J. A.: "Electromagnetic Theory" (McGraw-Hill, New York, 1941), Chapter VIII.

- (3) CAMPBELL, G. A., and FOSTER, R. M.: "Fourier Integrals" (D. Van Nostrand, New York, 1951), p. 90; pairs 751 and 752.
- (4) WATSON, G. N.: "Theory of Bessel Functions" (University Press, Cambridge, 1952), p. 229.
- (5) DOLPH, C. L.: "A Current Distribution for Broadside Arrays which Optimizes the Relationship between Beam Width and Side Lobe Level." *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 338.
- (6) VAN WIJNGAARDEN, A., and SCHEEN, W. L.: "Table of Fresnel Integrals," Report R 49 of the Computation Department of the Mathematical Centre at Amsterdam.
- (7) SILVER, S.: "Microwave Antenna Theory and Design" (McGraw-Hill, 1949), p. 196.
- (8) CAMPBELL, G. A., and FOSTER, R. M.: Reference 3, p. 86; pair 726.2.

(9) APPENDIX

To evaluate $J_1 = \int_{-\infty}^{\infty} \frac{\sin \frac{\pi \zeta}{a}}{\frac{\pi \zeta}{a}} e^{jD(\zeta - \psi)^2/2k} d\zeta$

$$J_1 = \frac{a}{j2\pi} e^{jD\psi^2/2k} \int_{-\infty}^{\infty} \frac{e^{jD\zeta^2/2k}}{\zeta} (\epsilon^{j\beta_1\zeta} - \epsilon^{j\beta_2\zeta}) d\zeta \quad (18)$$

where $\beta_1 = \frac{D}{k} \left(\frac{\pi k}{Da} - \psi \right)$ and $\beta_2 = -\frac{D}{k} \left(\frac{\pi k}{Da} + \psi \right)$

Consider $J_2 = \int_{-\infty}^{\infty} e^{j \left(\frac{D\zeta^2}{2k} + \beta\zeta \right)} d\zeta \quad (19)$

This is a Fourier transform with respect to β . From Reference 8

$$J_2 = -(1-j)\pi [C(\beta y_1) - jS(\beta y_1) + \frac{1}{2}(1-j)] \quad (20)$$

where $y_1 = \left(\frac{k}{D\pi} \right)^{1/2}$

Substituting eqn. (20) in eqn. (19) and then in eqn. (18) gives

$$J_1 = \frac{a}{2} (1+j) e^{jD\psi^2/2k} \{ C(\beta_1 y_1) - C(\beta_2 y_1) - j[S(\beta_1 y_1) - S(\beta_2 y_1)] \}$$

To evaluate $J_3 = \int_{-a}^a \cos \frac{\pi \zeta}{2a} e^{jD(\psi - \zeta)^2/2k} d\zeta$

$$J_3 = \frac{1}{2} e^{jD\psi^2/2k} \int_{-a}^a e^{jD\zeta^2/2k} (\epsilon^{j\beta_3\zeta} + \epsilon^{j\beta_4\zeta}) d\zeta \quad (21)$$

where $\beta_3 = \left(\frac{\pi}{2a} - \frac{D\psi}{k} \right)$ and $\beta_4 = -\left(\frac{\pi}{2a} + \frac{D\psi}{k} \right)$

Consider $J_4 = \int_{-a}^a e^{j \left(\frac{D\zeta^2}{2k} + \beta\zeta \right)} d\zeta$

$$J_4 = \frac{e^{-j\beta^2 k/2D}}{y_2} \{ C(y_1\beta + y_2a) - C(y_1\beta - y_2a) + j[S(y_1\beta + y_2a) - S(y_1\beta - y_2a)] \} \quad (22)$$

where $y_2 = \left(\frac{D}{\pi k} \right)^{1/2}$

Substituting eqn. (22) in eqn. (21)

$$J_3 = \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_3^2 k}{2D}\right)}}{2y_2} \left\{ C(y_1\beta_3 + y_2a) - C(y_1\beta_3 - y_2a) \right. \\ \left. + j[S(y_1\beta_3 + y_2a) - S(y_1\beta_3 - y_2a)] \right\} \\ + \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_4^2 k}{2D}\right)}}{2y_2} \left\{ C(y_1\beta_4 + y_2a) - C(y_1\beta_4 - y_2a) \right. \\ \left. + j[S(y_1\beta_4 + y_2a) - S(y_1\beta_4 - y_2a)] \right\} \\ \dots$$

To evaluate $J_5 = - \int_{-\infty}^{-a} \sin \frac{\pi \zeta}{a} e^{jD(\psi - \zeta)^2/2k} d\zeta$

and $J_6 = \int_a^{\infty} \sin \frac{\pi \zeta}{a} e^{jD(\psi - \zeta)^2/2k} d\zeta$

By procedure similar to the evaluation of J_3 ,

$$J_5 = \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_5^2 k}{2D}\right)}}{2y_2} \left[\frac{1+j}{2} + C(y_1\beta_5 - y_2a) + jS(y_1\beta_5 - y_2a) \right] \\ - \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_6^2 k}{2D}\right)}}{2y_2} \left[\frac{1+j}{2} + C(y_1\beta_5 - y_2a) + jS(y_1\beta_5 - y_2a) \right]$$

$$J_6 = \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_5^2 k}{2D}\right)}}{2y_2} \left[\frac{1+j}{2} - C(y_1\beta_5 + y_2a) - jS(y_1\beta_5 + y_2a) \right] \\ - \frac{e^{j\left(\frac{D\psi^2}{2k} - \frac{\beta_6^2 k}{2D}\right)}}{2y_2} \left[\frac{1+j}{2} - C(y_1\beta_6 + y_2a) - jS(y_1\beta_6 + y_2a) \right]$$

where $\beta_5 = \left(\frac{\pi}{a} - \frac{D\psi}{k}\right)$ and $\beta_6 = -\left(\frac{\pi}{a} + \frac{D\psi}{k}\right)$

POSSIBLE ERRORS OF A PARTICULAR WIDE-APERTURE DIRECTION-FINDER

By W. C. BAIN, M.A., B.Sc., Ph.D.

(The paper was first received 1st November, and in revised form 11th December, 1955. It was published as an INSTITUTION MONOGRAPH March, 1956.)

SUMMARY

A study is made of the extent to which the high-frequency cyclical direction-finder devised by Earp and Godfrey is subject to three common types of error. First, the result is presented of theoretical computations on the errors given by a practical fixed-aerial system in wave-interference conditions. The variance of these errors as the phase difference between the incident rays is varied is shown to be the same as that given by the ideal rotating-aerial version of this direction-finder; it is appreciably smaller than that of an Adcock for a system aperture of 4λ , and, in general, much smaller for an aperture of 10λ . Secondly, errors due to aerial interaction are considered for a system of 24 unipoles, 9 m high and 1.1 m in diameter, placed on a circle of 100 m diameter. It is concluded that the errors in indicated bearing are likely to be negligible, provided that the aerials not in use at any instant are terminated by a resistor equal to their nominal characteristic impedance. Finally, the polarization error is examined for such a system in which surface feeders are used, without an earth mat, on ground of moderate conductivity; it is found to be small, being greatest at the low-frequency end of the band, where the standard-wave error is about 1° at 3 Mc/s. With extended feeders the errors at low frequencies are reduced considerably.

LIST OF PRINCIPAL SYMBOLS

General Symbols.

- d = Diameter of aerial system.
 n = Total number of aerials in the system.
 λ = Wavelength of the radio waves being received.
 k = A number referring to a particular aerial at a bearing of $2k\pi/n$ from the centre of the system.
 ϕ_k = Phase of the signal at the k th aerial.
 ϵ_k = An error in the phase of the signal at the k th aerial.
 θ = Bearing of the incident radiation. (If two rays are being received, θ is taken to be the bearing of the stronger ray.)
 γ = The bearing indicated by the system.
 η = Error in indicated bearing (measured with respect to θ); hence $\eta = \gamma - \theta$.
 $\beta = \frac{\pi d}{\lambda} \cos \delta$.

Symbols used to describe two Incident Rays.

- r = Amplitude ratio of second ray to first ray as picked up by the aerial. It will be assumed to be less than unity; hence the bearing of the first ray is θ .
 α = Bearing of second ray with respect to first ray.
 ϕ_1 = Angle of elevation of first ray.
 ϕ_2 = Angle of elevation of second ray.
 β = Phase of second ray relative to that of first ray at the centre of the system.

(1) INTRODUCTION

A cyclical wide-aperture direction-finding system for high frequencies was proposed by Earp and Godfrey¹ in 1947. Wave-interference errors in an idealized form of this direction-finder

were considered in a previous paper.² Since the result of this investigation was promising, it seemed desirable to extend this work to cover the effects of wave interference on a more practical form of the system, and also to try to assess certain of the other errors to which an actual system would be liable. The errors studied were polarization errors and those due to aerial interaction, and they were investigated both experimentally and theoretically.

The ideal form of this direction-finder consists of a single aerial moving in a circle, the bearing being determined by finding the phase of the phase modulation imposed on the received signal by the circular movement. As is described by Earp and Godfrey, this system can be simulated in practice by the use of a number of fixed aerials spaced at equal intervals round the circle. In the paper the number of aerials will frequently be taken to be 24, and the diameter of the circle 100 m. The phase discriminator used in the equipment will be assumed to be linear, and to be able to measure phase differences in the range from -180° to $+180^\circ$.

(2) WAVE-INTERFERENCE ERRORS

Probably the most serious of the errors affecting narrow-aperture direction-finders are those caused by wave interference between rays which have travelled by different paths from the transmitter to the receiver. This Section deals with the effect of such conditions on cyclical systems, both of the idealized type and with fixed aerials.

The evaluation of wave-interference errors on systems with finite numbers of aerials is a very tedious process; the amount of effort involved in an extensive calculation would be prohibitive. Fortunately, the opportunity arose of having some of these calculations performed by an electronic computer, the Ace at the National Physical Laboratory. The results obtained are described below and are compared with those for other direction-finding systems.

(2.1) Problems Submitted to Computation

The case considered is that of two rays incident upon a direction-finding system, which can be an ideal cyclical system, a fixed-aerial cyclical system or a narrow-aperture Adcock. Instrumental errors are taken to be zero. Particular numerical values have been selected for the parameters of the rays, and a computation has been made of the mean square error averaged over all phase differences between the two incident rays, the true bearing being taken to be that of the stronger of the two. The results which follow give the error variance, σ_θ^2 , calculated from

$$\sigma_\theta^2 = \frac{1}{2\pi} \int_0^{2\pi} \eta^2 d\phi$$

This error, η , for the fixed-aerial cyclical system was computed on the Ace from the formula

$$\eta = \gamma, \text{ taking } \theta = 0$$

$$= \arg \left[\sum_{k=0}^{n-1} \psi_k \exp \left(\frac{2k\pi j}{n} \right) \right]$$

Correspondence on Monographs is invited for consideration with a view to publication.
 The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

as given in a previous paper.³ The method of deriving ψ_k is explained in Section 8.1. This formula for η applies to the system whether "phase compression"* is used or not, although it is assumed that the system is operated with sufficient phase compression to prevent the discriminator attempting to measure phase differences over 180° in magnitude. The errors for the ideal cyclical system and the Adcock system were calculated from formulae which have also been given previously by the author;² the cyclical-system errors can only be fixed between two limits in this way. The Adcock errors have been omitted from the Tables when the formula for their computation is unreliable.

* The technique of phase compression, proposed by Earp and Godfrey,¹ involves the use of higher than first-order differences in phase between successive aerials in the ring.

System apertures of 1, 4 and 10 λ have been used in the calculations, and the number of aerials has been taken to be 36. The angles of elevation chosen are approximately those for 1E, 1F and 2F reflections on a 700 km transmission path. However, the results for $\delta_1 = \delta_2 = 33^\circ$ can be taken as applying roughly to a case where both rays have angles of elevation under 15°, i.e. to a long-range transmission.

(2.2) Results

The following Tables give the error variance for each system in degrees squared. The parameters r , δ_1 and n have the same values throughout, these being $r = 0.8$, $\delta_1 = 33^\circ$ and $n = 36$.

Table 1
ERROR VARIANCE FOR $\delta_2 = 17^\circ$

α	Adcock system	$d/\lambda = 1$		$d/\lambda = 4$		$d/\lambda = 10$	
		Cyclical system		Cyclical system		Cyclical system	
		Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed-aerial
deg		deg ²	deg ²	deg ²	deg ²	deg ²	deg ²
1	—	0.41–2.78	1.61	0.19–0.32	0.27	0.001–0.001	0.001
2	—	1.65–10.4	6.23	0.75–1.23	1.01	0.003–0.003	0.004
3	—	3.74–21.7	13.28	1.60–2.55	2.11	0.013–0.014	0.014
4	—	6.67–35.5	22.04	2.66–4.04	3.38	0.045–0.048	0.047
5	—	10.1–53.4	31.84	3.71–5.52	4.65	0.117–0.125	0.121
6	—	14.5–70.9	42.13	4.72–6.76	5.79	0.223–0.241	0.235

Table 2
ERROR VARIANCE FOR $\delta_2 = 33^\circ$

α	Adcock system	$d/\lambda = 1$		$d/\lambda = 4$		$d/\lambda = 10$	
		Cyclical system		Cyclical system		Cyclical system	
		Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed aerial
deg	deg ²	deg ²	deg ²	deg ²	deg ²	deg ²	deg ²
1	0.89	0.87–0.89	0.88	0.80–0.81	0.81	0.59–0.60	0.60
2	3.56	3.40–3.50	3.43	2.62–2.68	2.65	1.40–1.41	1.41
3	8.00	7.23–7.66	7.41	4.71–4.85	4.79	1.91–1.93	1.94
4	14.2	11.9–13.2	12.51	6.76–7.00	6.88	2.07–2.09	2.09
5	22.2	17.2–19.7	18.46	8.55–8.89	8.74	1.82–1.85	1.83
6	32.0	22.8–27.2	24.99	10.1–10.7	10.31	1.20–1.22	1.22

Table 3
ERROR VARIANCE FOR $\delta_2 = 53^\circ$

α	Adcock system	$d/\lambda = 1$		$d/\lambda = 4$		$d/\lambda = 10$	
		Cyclical system		Cyclical system		Cyclical system	
		Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed-aerial	Ideal (limits)	Fixed-aerial
deg	deg ²	deg ²	deg ²	deg ²	deg ²	deg ²	deg ²
1	0.25	0.09–0.97	0.16	0.008–0.013	0.011	2.5–2.6 $\times 10^{-4}$	0.006
2	1.00	0.38–3.80	0.64	0.032–0.047	0.042	0.0012–0.0013	0.001
3	2.25	0.86–8.33	1.43	0.07–0.10	0.08	0.0028–0.0029	0.010
4	4.00	1.49–14.7	2.53	0.10–0.16	0.15	0.0056–0.0060	0.006
5	6.25	2.36–22.4	3.92	0.15–0.22	0.20	0.010–0.011	0.017
6	8.99	3.46–31.0	5.60	0.20–0.28	0.25	0.017–0.019	0.018

A 20-aerial system was also examined in all the cases for $\alpha = 1^\circ$ and 6° . The error variances obtained nowhere differed from the corresponding ones for the 36-aerial system by more than 0.04 deg^2 . In addition, tests were made to discover whether the value of θ affected the results for the 36-aerial system, by trying values of θ of $1^\circ, 2^\circ \dots 6^\circ$; no appreciable change was found. A fact related to this was also discovered, namely that there were no errors for $\alpha = 0$ in the 36-aerial system, as is already known for the ideal and the Adcock systems.

(2.3) Discussion of the Results

The most striking feature of the results is the close similarity between those for the ideal and practical systems, at least in those cases where the ideal system performance can be assessed fairly accurately. It appears that the use of a finite number of aeri-als does not seriously upset the behaviour of the system, provided, of course, that the phase discriminator is not set the ask of measuring phase differences exceeding 180° . Even the 20-aerial system would appear to work satisfactorily throughout $10:1$ frequency range.

The improvement given by both wide-aperture systems upon the narrow-aperture Adcock is quite considerable when $d/\lambda = 4$ to 10 . It may be mentioned here that the Adcock values for $\delta_1 = 17^\circ$, which are not quoted, are in all cases greater than those for $\delta_2 = 33^\circ$.

(3) EFFECT OF AERIAL INTERACTION

In the experiments described here the aerial used was a vertical dipole of cylindrical shape, 9 m in height and 1.1 m in diameter, and tapered towards the base. It was constructed in cage-form from 16 s.w.g. wires, joined together electrically in the horizontal plane at five different levels by means of aluminium strip. Three such aeri-als were made, one being portable and two fixed.

The main objects were to assess the magnitude of errors likely to arise from aerial interaction, and to find the best way of terminating those aeri-als not in use at any moment. It will be appreciated that in the cyclical direction-finder at a given instant only one aerial is connected to a receiver; the comparison between signals at different aeri-als in the system is achieved by means of delay techniques. The termination of the aeri-als not in use is therefore not determined directly by the operation of the system and may be selected to minimize aerial interaction effects. The methods of termination which are studied here are

- (a) A short-circuit.
- (b) An open-circuit.
- (c) A 100-ohm resistor, this being equal to the nominal characteristic impedance of the aerial.

(3.1) Experimental Arrangements

The experimental work was confined to assessing the effect on one aerial, C, produced by another aerial, A, situated at a distance equal to the spacing of two elements in the direction-finding system; the effect of all the other aeri-als in the ring is considered later theoretically. If a plane wave is incident on these two aeri-als, and if aerial A is moved round aerial C (see Fig. 1), the signal picked up by C will vary in phase as shown in eqn. (6). Several of the parameters in this equation are of unknown magnitude, but it can be seen that, if the phase and amplitude of the signal in aerial C are taken as the polar co-ordinates of a point, the locus of this point produced by the movement of aerial C will be a circle. The size of this circle will depend on the magnitude of the aerial interaction.

The aeri-als were placed as shown in Fig. 1; R (an aerial giving reference signal) and C were fixed wide-band aeri-als, A was a portable wide-band aerial, and T was an ordinary thin unipole energized by a low-power transmitter. An earth mat was

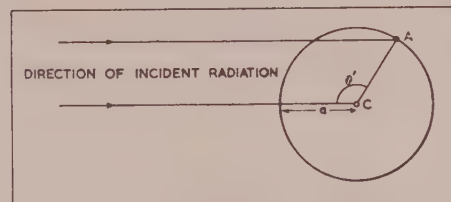
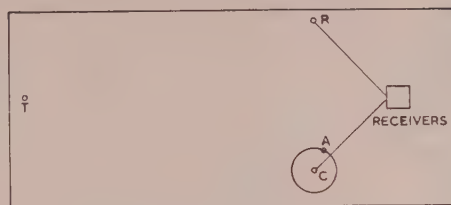


Fig. 1.—Diagrams of the experimental arrangements.

- T. Transmitting aerial.
- R. Reference aerial (fixed).
- C. System aerial (fixed).
- A. System aerial (movable).
- TR = TC = 300 m.
- AC = 10.5 m.

employed, of circular shape with its centre at the point C; it had a radius of 15 m and a mesh size 0.61 m square, thus being of very similar construction to the earth mat often used with an h.f. Adcock installation. It was earthed at the centre, and earth stakes were also inserted at 10° intervals on a circle of radius 10.5 m, at the points where the portable aeri-als were to be placed. Additional earths were also inserted at 18° intervals round the circumference of the circle. The wire used in the mat construction was of 16 s.w.g., and every cross-over point was soldered. The plane of the mat was made horizontal.

From the central system aerial C and the reference aerial R feeders were run to a trailer containing the necessary amplitude- and phase-measuring equipment. Feeder of 70-ohm impedance was employed, with a resistive matching unit at the aerial base. Each input fed a cathode-follower, and the grids of these valves could be connected together by means of a switch for lining-up purposes. Each cathode-follower output was connected to one channel of a twin-channel receiver of the Admiralty type FHB. An output was taken from each channel to phase-measuring equipment of the sum-and-difference type similar to that used by Ross, Bramley and Ashwell.⁴

Suitable electrical alignment facilities were provided, so that there were no errors from differential effects between the channels in the circuits following the line-up unit. It must be assumed that the aeri-als and feeders did not change in their electrical properties during a cycle of rotation of the portable aerial. It is not essential that the feeders should be identical, provided that the method is applied only to results at one frequency; comparisons between phases on different frequencies, for instance, cannot be made directly from the results, for the difference in electrical length of the feeders is not the same in each case.

(3.2) Results

Measurements were made of the impedance of an aerial, with the results given in Fig. 2. It will be seen that the variation of impedance over the band is not great, except at the low-frequency end.

Experiments in moving a portable aerial around the central system aerial were carried out with frequencies at 3 Mc/s intervals

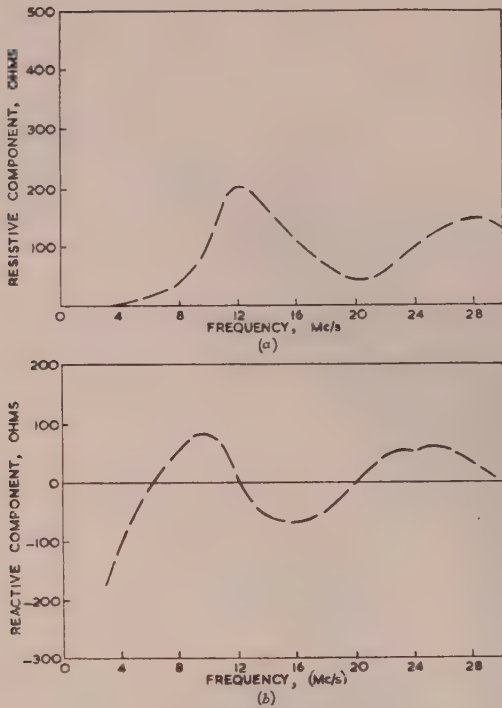


Fig. 2.—Impedance of a wide-band aerial.

(a) Resistive component.
(b) Reactive component.

from 3 to 30 Mc/s. It was found best to keep the frequency constant while the portable aerial was moved round the whole circle, its termination being changed between all three values at each position. In this way random errors due to changes in local conditions were kept as low as possible.

The phase and amplitude results of these experiments were plotted as polar co-ordinates separately for each frequency and aerial termination, as explained in Section 3.1. Fig. 3 shows a

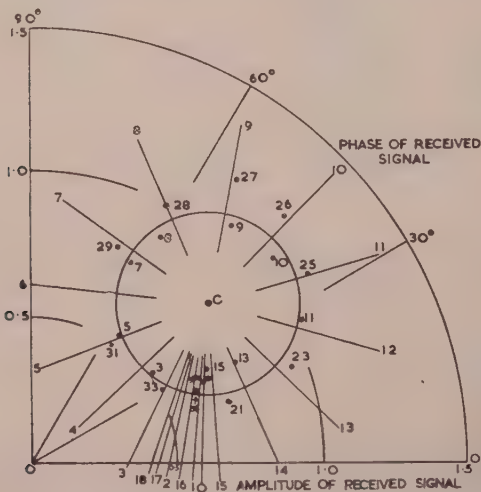


Fig. 3.—Aerial interaction locus for 15 Mc/s, with open-circuit termination of the portable aerial.

Observed points are shown with the best theoretical circle fitted to them.

typical result, with a theoretical circle fitted to the points by a least-squares method. On this diagram each radial line crosses the circle at the point where the observed point with the same number should appear. The number shown with each point and line, when multiplied by 10, gives the bearing of the movable aerial. The lines are numbered only from 0 to 18, and a point with a number, say m , greater than 18 should lie on the line marked $(36 - m)$. The interaction ratio—the ratio of the radius of this circle to the distance of its centre from the origin—has also been calculated, and a plot of its values against frequency

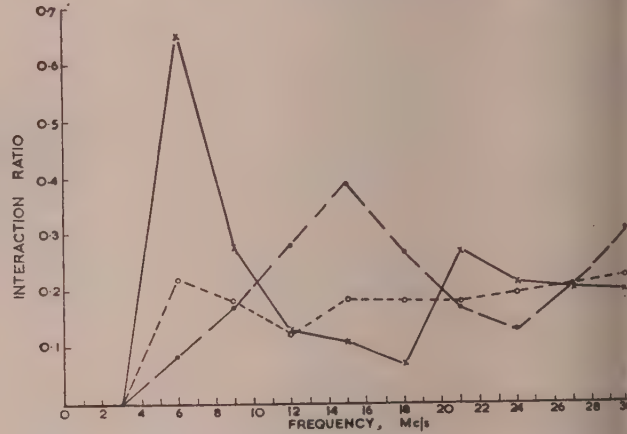


Fig. 4.—The interaction ratio from the aerial interaction loci.

Termination of portable aerial:

- × Short circuit.
- Open circuit.
- 100-ohm resistor.

is shown in Fig. 4 for various types of termination of the movable aerial.

A value was derived in each case for the standard deviation of the interaction ratio. This was found not to exceed 0.023 for any frequency or type of termination. It may therefore be considered that the points follow reasonably well the theoretical distribution.

The interaction ratio shows a large maximum near the $\lambda/4$ resonance of the aerial when the portable aerial is short-circuited to earth. There is another maximum in the open-circuit case near 15 Mc/s. The low values at the low-frequency end of the spectrum for all terminations can be derived from the theory of the system. As shown in eqn. (7), the interaction ratio is proportional to the ratio $Z_1/(Z_S + Z_K)$, where Z_S is the self-impedance of an aerial, Z_1 is the mutual impedance between the two aerials, and Z_K is the termination of the movable aerial. Now at frequencies below that of the first aerial resonance the self-impedance increases rapidly (becoming mainly capacitive) as the frequency decreases, but the mutual impedance decreases somewhat. Hence the interaction ratio will decrease rapidly with the frequency. This argument requires some elaboration when the termination is an open-circuit, but the result is the same. The ratio is remarkably constant for a 100-ohm resistive termination of the portable aerial, and is not very high.

(3.3) Bearing Errors

If there are errors, e_k , in the phase of the signal at the k th aerial, the resultant bearing error η can be obtained from the formula

$$\eta = -\frac{2}{n\beta} \sum_{k=0}^{n-1} e_k \sin \left(\theta - \frac{2k\pi}{n} \right) \dots \quad (10)$$

provided that the phase errors are small compared with a half-cycle (see Section 8.3). The phase errors described in previous sections were due to a single neighbouring aerial, but here a more general expression will be used, including the effect of all aials in the ring; it is derived in Section 8.4.1.

In Section 8.4.2 the bearing error due to these phase errors is shown to be

$$= -\frac{2n}{\beta^2} \sum_{s=1}^{\infty} s \sin(s n \theta) \sum_{l=1}^{n/2} (2 - \delta_{l, n/2}) (-1)^s r_l \cos \phi_l J_{sn} \left(2\beta \sin \frac{l\pi}{n} \right) \quad (30)$$

for an even number of aerals. The restriction on the validity of this expression is that R should be small, when

$$R e^{j\alpha} = \sum_{l=1}^{n-1} r_l e^{j\theta_{kl}} \quad (16)$$

where r_l is defined with eqn. (14) and θ_{kl} with eqn. (15).

Eqn. (30) gives an error which vanishes when θ is any multiple of π/n , corresponding to the planes of symmetry of the aerial system. Also, because of the properties of the Bessel functions involved, the error will rapidly fall to a negligible amount as n is increased above 2β , i.e. as the spacing between adjacent aerals is reduced below $\lambda/2$. Hence we reach the conclusion that for a aerial system of given diameter the errors due to aerial interaction are reduced to negligible proportions by increasing the number of aerals until the aerial spacing is less than $\lambda/2$, provided that the values of r_l do not increase too much. This result is, of course, quite independent of the number of stages of phase compression used.

To apply eqn. (30) to an actual system it is necessary to know the values of r_l and ϕ_l . This is probably best done by experiment on a complete system; if the resistive termination of aerals not in use is employed, an e.m.f. could be injected at the base of aerial No. 0, and $r_l e^{j\theta_{kl}}$ could be taken to be the ratio of the base currents in the l th and 0th aerals.

If this cannot be done, the values must be found by calculations from the results of experiments such as those described in Section 3.4. For a 10.5 m aerial spacing, for instance, the interaction ratio was found to be 0.2. This can be taken as giving roughly the value of r_l for this spacing. In the following it will be assumed that r_l is inversely proportional to the distance between aerals subtending an angle of $2l\pi/n$ rad at the centre of the circle; this should be very roughly true, for r_l is approximately proportional to the mutual impedance between such aerals if interaction is not too great.

On this basis the bearing error for a 24-aerial 100 m-diameter system can be estimated: r_l can be taken to be given by

$$r_l = \frac{0.021}{\left| \sin \frac{l\pi}{n} \right|}$$

The validity of the assumption about the magnitude of R can now be checked. From eqn. (16),

$$R^2 = \sum_l \sum_i r_l r_i \cos(\theta_{kl} - \theta_{ki})$$

It is now assumed that the values of θ_{kl} are uncorrelated for different values of l ; this is not unreasonable, since each θ_{kl} contains the term ϕ_l , which depends in a complicated way upon mutual impedances. The expression for R^2 therefore reduces to

$$R^2 = \sum_{l=1}^{n-1} r_l^2$$

for a large number of aerals. This gives $R = 0.37$; hence eqn. (30) can be used without a great risk of error.

To obtain the bearing error when the aerial spacing is always somewhat less than λ , only the term in $s = 1$ in eqn. (30) need be considered. The value of $|\eta|$ at several frequencies has been computed, with the results given in Table 4. $(-1)^l \cos \phi_l$ has

Table 4

UPPER LIMITS TO BEARING ERROR DUE TO AERIAL INTERACTION ON THE 24-AERIAL SYSTEM

Frequency, Mc/s	9.00	10.50	11.25	12.00	12.75	13.50	15.00
Aperture, λ	3.00	3.50	3.75	4.00	4.25	4.50	5.00
Bearing error, deg	0.03	0.18	0.33	0.49	0.62	0.60	0.39

been assumed to be 1 for all values of l , and $\sin n\theta$ has also been taken as 1, to give an upper limit to the error. On the basis of the assumption made earlier that ϕ_l was random, the true error will be appreciably less than this.

In this system the distance between aerals exceeds $\lambda/2$ when $d/\lambda = 3.83$, i.e. at 11.5 Mc/s. It can be seen that the error does, in fact, fall off rapidly below this frequency and becomes negligible below 10 Mc/s. The fall in the error as the frequency is increased above 13 Mc/s is due to the $1/\beta^2$ factor in eqn. (30). These features of the bearing error will still be present even if the values of r_l are changed considerably, provided that they are not made too large. It will also be noted that, for values of r_l of the magnitude quoted here, even the greatest of the errors is very small.

(4) POLARIZATION ERRORS

The aerals used in the polarization-error experiments were the same as those employed in the work on aerial interaction. The object of the investigation was to find the polarization error due to the feeders connected to the aerals rather than that due to extraneous objects which happened to be on the site selected. The possibility of reducing the error by continuing the feeders to some distance beyond the aerals was also studied.

(4.1) Experimental Methods

A method was adopted which would give the polarization error in the phase at a single aerial due to the presence of its own feeder in such a way that effects due to the site were eliminated. A site was selected as far removed from obstacles as possible, and the two aerals were placed in line with the mast carrying the transmitter. The trailer containing the receiving and phase-measuring equipment was also placed in this line, the feeders being run directly from the aerals to the trailer [see Fig. 5(a)]. Aerial A was treated purely as a reference aerial, whereas aerial B was regarded as one of the system aerals and the d.f. system centre was assumed to lie at a point on a circle of 50 m radius around B. A dummy feeder, whose inner conductor was left unconnected, was run from the earth terminal at the base of aerial B to the system centre [see Fig. 5(b)]. The diameter of its screen was 0.95 cm. The dummy feeder then simulated the feeder running from the system centre to an aerial in the actual direction-finding system. The other feeder, running direct from aerial B to the trailer, was, of course, required to convey the signals picked up by aerial B to the receiving equipment; it should not affect the results, since the horizontal component of the electric field is at right angles to it. The angle of elevation of the transmitter from aerial B was 10.5° , and the distance separating them in the horizontal plane was 138 m.

The principle of obtaining the polarization error was to measure the phase difference and the amplitude ratio of the signals for cases where the transmitter loop was vertical or at $\pm 84^\circ$ to the vertical. The dummy feeder was then transferred

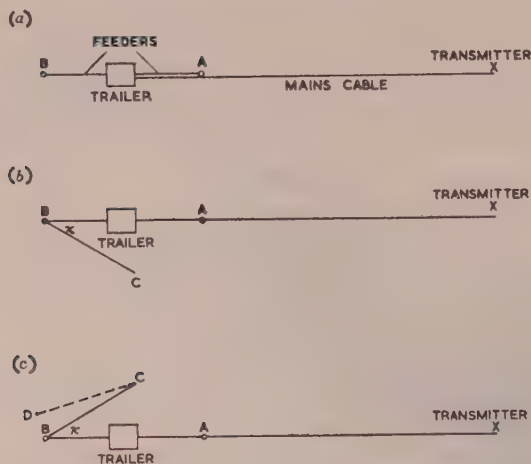


Fig. 5.—Arrangement of the site in the polarization-error experiments.

The transmitter is at the top of a 100 ft mast.

(a) Layout of cables and feeders carrying signals to the receivers.

(b) and (c). Arrangements of dummy feeder to obtain one measurement of polarization error in the phase. The broken line CD in (c) shows the position which would be occupied by the feeder to a neighbouring aerial if the whole ring of aerials were present.

to the other side of the line joining the aerials, as shown in Fig. 5(c), and the measurements were repeated. By subtracting one set of results from the other, the polarization error due to the presence of the dummy feeder could be obtained.

The method of calculating the error is explained more precisely in Section 8.5. Two independent expressions, eqns. (38) and (39), are derived for the in-phase polarization error e_p , and two other expressions, eqns. (40) and (41), for the quadrature error e_q . These two errors, of course, are errors not in the bearing but in the phase. The expression "in-phase error" is used here to signify an error which would appear in a measured phase, whereas "quadrature error" is used for an error which appears as an alteration in signal amplitude, but which would become a phase error if the phase of the horizontal component of incident electric field were changed by 90° . The means of the results obtained from each of these pairs of values were taken as the best values of e_p and e_q . The total polarization error, e_T , in the phase difference was obtained from the usual relation

$$e_T = \sqrt{(e_p^2 + e_q^2)}$$

The experiments were carried out for a number of pairs of values of $\chi = \pm \chi_0$, where χ is the angle between the direction

of the dummy feeder and BA, the measurements for the two values $\pm \chi_0$ being carried out with as small a time interval as possible. The transmitter covered the frequency band 3–20 Mc/s in three ranges, which were used in turn.

The arrangement of apparatus in the trailer was as given in Section 3.3. Measurements at each frequency were carried out by first setting the transmitter loop to be vertical and tuning in the signal. Lining-up and measurement of phase and amplitude were performed as described in Section 3.3, and this process was then repeated for loop tilts of $\pm 84^\circ$ to the vertical.

To simulate as closely as possible the layout of the actual direction-finding system, the outer of the dummy feeder should be connected to earth at its far end, which represents the system centre. This was done in the first of the experiments, but it was discovered that the results were quite unaffected by disconnecting the earth. In subsequent work the end of the dummy feeder was left disconnected to save the labour of driving in more earth stakes. It may be noted here that earths were provided at the base of each aerial, consisting of four earth stakes driven to depth of about 2 m.

(4.2) Results

The polarization-error results which follow are all expressed in terms of a loop tilt of 84° and an angle of elevation of 10.5° . The errors quoted here are therefore converted to the standard wave error by multiplying by the factor 0.8, provided that a perfectly conducting earth is assumed. As the measured conductivity found on the site was about 7×10^{-3} mho/m, this derivation of standard wave error does not apply strictly, of course, but it may still serve as a useful indication of the magnitude of the effect.

Now suppose the transmitter was at a bearing of 0° from the system aerial. The polarization error in the phase was measured in the frequency range 3–20 Mc/s for the positions of the dummy feeder given by $\chi_0 = 30^\circ, 60^\circ, 90^\circ, 120^\circ$ and 150° . The results for the total error are given in Fig. 6, and a specimen of the individual results for the in-phase and quadrature errors is given in Fig. 7.

The total errors in the phase are not very high, except at the top of the frequency range, where there is a noticeable increase of error with the frequency. An examination of the curves such as that shown in Fig. 7 reveals that the increase is mainly in the in-phase component of error. At the low-frequency end of the range the quadrature component is predominant, and is nearly constant with frequency at a value of about 4° for $\chi = \pm 90^\circ$.

It is important to know the functional dependence of the error

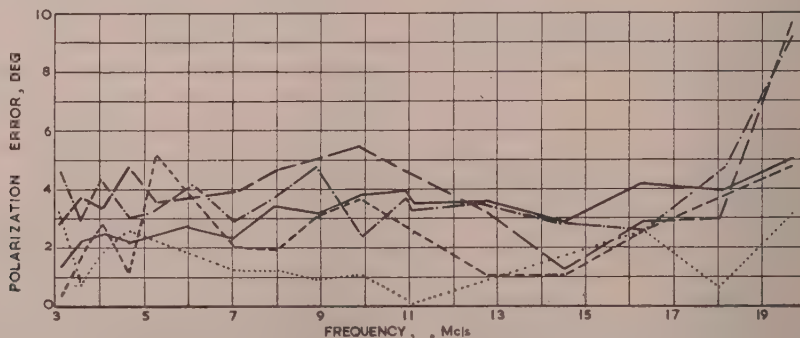


Fig. 6.—Total polarization error in the phase for different positions, χ_0 , of the dummy feeder.

— 30°
 - - - 60°
 90°
 - · - · - 120°
 - - - - - 150°

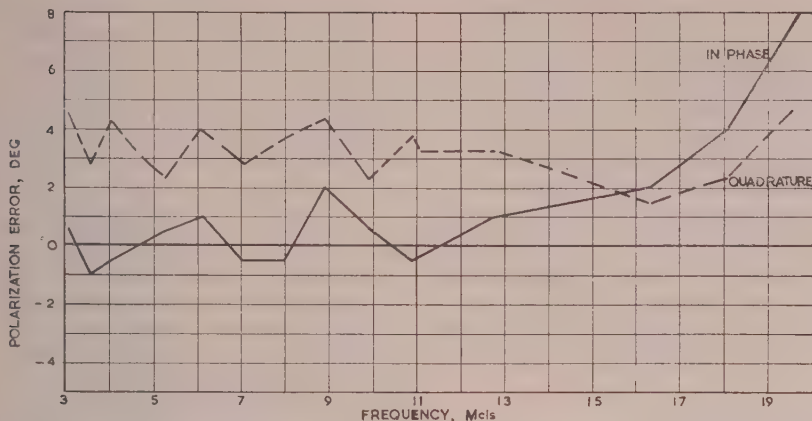


Fig. 7.—In-phase and quadrature polarization errors in the phase for the positions of the dummy feeder given by $\chi_0 = 90^\circ$. The in-phase error occurs as a phase advance where shown positive.

on the position of the dummy feeder (χ), for in the complete section-finding system the feeders to the different aerials will be inclined at various angles to the direction of propagation. This dependence was checked in detail for the frequency of 10 Mc/s; the error curve was found to be approximately sinu-

$+ 180^\circ$. The polarization error with extended feeder was then derived by combining these results in the usual way (see Section 8.5) with those for feeders at $\chi = -\chi_0$ and $\chi = -(\chi_0 + 180^\circ)$.

Results of this type are shown in Fig. 8 for $\chi_0 = 60^\circ$. It can be seen that the presence of the extended feeder reduces the error

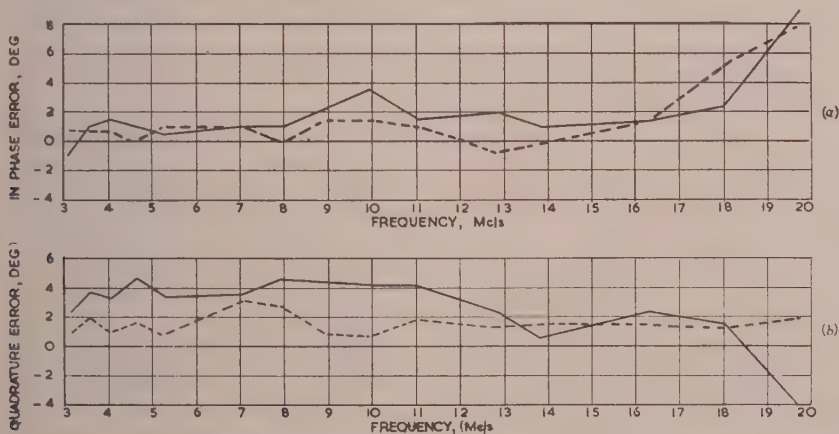


Fig. 8.—Polarization errors in the phase for the system with and without extended feeders; $\chi_0 = 60^\circ$.

— No extended feeder.
- - - Extended feeder.

oidal with zeros at 0° and 180° , as might be expected theoretically.

A few measurements of polarization error were made at frequencies above 20 Mc/s with the value of 90° for χ_0 . The total error in the phase was found to rise to about 13° at 24 Mc/s, and then to fall off again at higher frequencies. A possible explanation of this phenomenon is that the polar diagram of the aerials used probably has a minimum near the angle of elevation of 15° at a frequency of 24 Mc/s.

It has been suggested that the polarization errors in this system could be reduced if dummy feeders were run from each aerial in the opposite direction to the system centre, provided that these feeders had the same external radius and were of the same length as the feeders carrying the signal to the receiving equipment. This condition was simulated here by placing, in addition to the dummy feeder at $\chi = \chi_0$, a second dummy feeder at $\chi = \chi_0$

appreciably over most of the frequency range, although there is little or no improvement in the region of 20 Mc/s. If $\chi_0 = 90^\circ$ it can, of course, be said on theoretical grounds that there is no error with an extended feeder. It therefore appears that below 15 Mc/s the extended feeders reduce the polarization error by a factor of 2, but above this frequency the benefit is doubtful unless the wave is incident in the plane normal to the feeders.

Some other observations were made which may help in determining the polarization error of wide-aperture systems. With the dummy feeder at $\chi = 90^\circ$ the transmitter loop was set at 84° to the vertical and at a frequency of 20 Mc/s. The feeder, which had an insulating outer covering, was then gradually coiled up, beginning with the end distant from the aerial. No change could be detected in the phase or amplitude of the signal picked up by the aerial until the length of the feeder had been reduced to 10 m. The feeder was again extended to the full 50 m length

and the distant end was held 2 m in the air, thus subjecting it to a much greater horizontal electric field than when it lay on the ground. Again no change was observed in the signal received. This seems to indicate that at 20 Mc/s signals are not propagated readily along the outer screen of a feeder lying on the ground.

Tests were also carried out in which the feeder carrying the signals from the system aerial to the trailer was run out 50 m from the aerial in the direction $\chi = 90^\circ$ or 270° and from there returned to the trailer direct by another feeder. It was found that the polarization error was the same as in the case of the dummy feeder.

Experiments were also performed on placing a feeder in the position which it would have occupied if connected to the neighbouring aerial in the direction-finding system [see Fig. 5(c)]. This was found to have no detectable effect on the error at any frequency.

(4.3) Bearing Errors

The polarization errors in bearing can be deduced from the errors in the phase; if in-phase errors in the phase are used, the resulting bearing polarization errors will be in-phase in the usual sense, and if the phase errors are total, the bearing errors will be total. The numerical data in this Section deal with total errors.

Suppose that the phase error in the k th aerial is e_k . The expression for the error η in the indicated bearing has already been given; if e_k is small it is

$$\eta = \frac{2}{n\beta} \sum e_k \sin \left(\frac{2k\pi}{n} - \theta \right) \quad (10)$$

It has previously been stated that the variation of phase error with the position of the dummy feeder is approximately sinusoidal, so to a first approximation e_k may be taken to be of the form

$$e_k = a \sin \left(\theta - \frac{2k\pi}{n} \right)$$

If this value is substituted in eqn. (10) we obtain

$$\eta = \pm a/\beta$$

The bearing error of a/β is not likely to be serious when we consider the values of a given by the measurements, for which the angle of elevation of the transmitter was 10.5° . These ranged from about 4° at 3 Mc/s to about 9° at 20 Mc/s, and would give bearing errors of 1.3° and 0.4° respectively. (The corresponding standard wave error at 3 Mc/s is 1.0° .) Above 3 Mc/s the error would fall off steadily to a value of 0.2° at 18 Mc/s, but above this frequency the error would rise again. The results obtained above 20 Mc/s suggest that the maximum value of the error at these frequencies would be 0.5° at 23.5 Mc/s. It should be noted that the total polarization error in the phase has been used to derive these figures.

(4.4) Discussion

The observations made here indicate that polarization errors need not be a serious problem on this wide-aperture system. Even with no earth mat or extended feeders, the errors remain small on ground of conductivity 7×10^{-3} mho/m. If extended feeders are used the largest of the errors which remain—namely those from 3 to 6 Mc/s—become negligible. The provision of an earth mat therefore does not appear essential to keep polarization errors small, provided that the ground is not of poor conductivity and that aerials of the type and dimensions described are used.

It is interesting to attempt to deduce from the results the polarization error to be expected on a U-Adcock direction-finder with surface (screened) feeders and no mat. Suppose that a

wave of frequency 3 Mc/s and with polarization the same as in the wide-aperture experiments is travelling from the north towards the Adcock aerials. In the east and west aerials there will be errors in the phase (in-phase or quadrature) of 4° in the opposite sense. This will induce a voltage proportional to $\sin 4^\circ$ in the east-west pair. In the north-south pair a voltage will be induced proportional to $\sin [(\pi d/\lambda) \cos \delta]$. Now, since $[(\pi d/\lambda) \cos \delta] \simeq 11^\circ$, the total polarization error will be 20° . After reduction by the factor already quoted this gives a standard wave error of 16° . The observed polarization error which compares most closely with the experimental conditions is that measured by Smith-Rose and Ross.⁵ They give a figure of 8 for a transmitter with an angle of elevation of 11° and a loop at 75° tilt; conversion of this by the usual process gives a standard wave error of 16° . Of course, too much weight should not be attached to the fortuitous agreement, especially since the Adcock aerials used by Smith-Rose and Ross were only 6 m high instead of the more usual 9 m. None the less, it shows that an effect of the correct order of magnitude can be predicted from the phase results.

The phase errors were practically constant with frequency up to 15 Mc/s, so that on this basis the Adcock polarization error should be inversely proportional to frequency up to 10 Mc/s at least. This does not agree well with the curves given by Smith-Rose and Ross, but, on the other hand, some earlier work of Barfield⁶ shows this effect very plainly.

The absence of change in the phase results with length of the dummy feeder (provided that this exceeds 10 m) implies that the polarization error of the cyclical system will be inversely proportional to the aperture at any given frequency. This may well be a characteristic of most, if not all, wide-aperture systems.

(5) CONCLUSIONS

The performance of the cyclical system considered here has been found to be identical for practical purposes whether 3 or 20 fixed aerials are used or a single continuously-rotating aerial under the same specified conditions of wave interference. Such a system gives a substantially lower error variance than the Adcock system if it has an aperture of 4λ , and is better still for an aperture of 10λ .

The aerial interaction experiments revealed that it was best to terminate the aerials not in use by a resistance equal to the nominal aerial impedance (100 ohms). In this case the presence of a single neighbouring aerial of the 24-aerial 100 m-diameter system could alter the phase at an aerial by 9° at the most. An examination of the combined effect of all the aerials in the ring indicated that the error in indicated bearing was likely to be negligible.

If the aerials are 9 m high, have surface feeders and are situated on ground of moderate conductivity (0.7×10^{-2} mho/m) polarization errors should be small even if no earth mat is used. The highest polarization error in bearing is expected to occur at the low-frequency end of the band and should be about 1.3° at 3 Mc/s for a transmitter with a loop tilt of 84° to the vertical and an angle of elevation of 10.5° . At frequencies above 4 Mc/s it should be less than 1° . With extended feeders the errors at low frequencies should be reduced substantially.

(6) ACKNOWLEDGMENTS

The author is indebted to the Director, National Physics Laboratory, for allowing the Ace to be made available for the computing of wave interference errors. The computations from the formulae given in the paper were carried out under the direction of Mr. E. J. York, who also gave helpful advice on the problems. The author desires to acknowledge the assistance

en by his colleagues Messrs. J. Bell, J. Meadows and R. E. hemel, who carried out much of the experimental work and mputations.

The work described above was carried out as part of the ogramme of the Radio Research Board. The paper is blished by permission of the Director of Radio Research of e Department of Scientific and Industrial Research.

(7) REFERENCES

- EARP, C. W., and GODFREY, R. M.: "Radio Direction-Finding by the Cyclical Differential Measurement of Phase," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 705.
- BAIN, W. C.: "The Calculation of Wave-Interference Errors on a Direction-Finder employing Cyclical Differential Measurement of Phase," *Proceedings I.E.E.*, Paper No. 1545 R, September, 1953 (**100**, Part III, p. 253).
- BAIN, W. C.: "The Theoretical Design of Direction-Finding Systems for High Frequencies," *ibid.*, Paper No. 1960 R, January, 1956 (**103** B, p. 113).
- ROSS, W., BRAMLEY, E. N., and ASHWELL, G. E.: "A Phase-Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves," *ibid.*, Paper No. 1134 R, July, 1951 (**98**, Part III, p. 294).
- SMITH-ROSE, R. L., and ROSS, W.: "The Use of Earth Mats to Reduce the Polarization Error of U-Adcock Direction-Finders," *Journal I.E.E.*, 1947, **94**, Part III, p. 91.
- BARFIELD, R. H.: "Some Principles underlying the Design of Spaced-Aerial Direction-Finders," *ibid.*, 1935, **76**, p. 423.

(8) APPENDICES

1) Phase of the Received Signals in Conditions of Wave Interference

As seen from the centre of the system, the aerials are separated an angle of $2\pi/n$. Suppose the voltage induced in the k th aerial is V_k . Then, if two rays are incident on the system,

$$V_k = \exp j \left[\omega t + \frac{\pi d}{\lambda} \cos \delta_1 \cos \left(\theta - \frac{2k\pi}{n} \right) \right] + r \exp j \left[\omega t + \phi + \frac{\pi d}{\lambda} \cos \delta_2 \cos \left(\theta + \alpha - \frac{2k\pi}{n} \right) \right] \quad (1)$$

$$\theta_{1k} = \frac{\pi d}{\lambda} \cos \delta_1 \cos \left(\theta - \frac{2k\pi}{n} \right)$$

$$\theta_{2k} = \frac{\pi d}{\lambda} \cos \delta_2 \cos \left(\theta + \alpha - \frac{2k\pi}{n} \right)$$

$$V_k = \exp j(\omega t + \theta_{1k}) [1 + r \exp j(\theta_{2k} - \theta_{1k} + \phi)]$$

The phase of this signal, ψ_k , is given by

$$\psi_k = \theta_{1k} + \arctan \frac{r \sin(\theta_{2k} - \theta_{1k} + \phi)}{1 + r \cos(\theta_{2k} - \theta_{1k} + \phi)} \quad (2)$$

(8.2) Interaction Effects between Two Aerials

Suppose the transmitter applies to the base terminal of its aerial, T, a voltage V . If this aerial has impedance Z_T , a current V/Z_T will flow in it, and this will cause a current I , say, to flow in the aerial C at its base terminal (see Fig. 1). Then, by the principle of reciprocity, if a voltage V were applied to the base terminal of the aerial C a current I would flow in the aerial T at its base terminal. If the current I can be calculated in this second case, the current in the original problem is known and thus the voltage in the load of the aerial C.

Suppose now that the two aerials C and A are terminated in loads Z_L and Z_K respectively. C is driven by a voltage V , since the inverse problem is being considered. The currents in the loads of the aerials are therefore given by

$$(Z_S + Z_L)I_1 + Z_1I_2 = V \quad (3)$$

$$Z_1I_1 + (Z_S + Z_K)I_2 = 0 \quad (4)$$

where Z_S, Z_1 are the self- and mutual impedances of aerials with the configuration of aerials in question, I_1 is the current in the driven aerial (C), and I_2 that in the other (A). The solution of eqns. (3) and (4) is

$$\left. \begin{aligned} I_1 &= \frac{V(Z_S + Z_K)}{(Z_S + Z_L)(Z_S + Z_K) - Z_1^2} \\ I_2 &= \frac{-VZ_1}{(Z_S + Z_L)(Z_S + Z_K) - Z_1^2} \end{aligned} \right\} \quad (5)$$

Now the e.m.f. at the base of aerial T will be proportional to the currents I_1 and I_2 and may therefore be written

$$g_1I_1 + g_2I_2 \exp(j2\pi a \cos \theta' / \lambda)$$

where $2\pi a \cos \theta' / \lambda$ represents the phase difference between signals arriving at T from A and C; g_1 and g_2 will depend upon the current distributions in aerials C and A and upon certain other geometrical quantities.

Hence the current at base of aerial T is

$$\frac{1}{Z_T} (g_1I_1 + g_2I_2 \exp(j2\pi a \cos \theta' / \lambda))$$

This is the same as the current in the original problem. The voltage across the load of aerial C in the original problem is therefore

$$\frac{Z_C V [g_1(Z_S + Z_K) - g_2Z_1 \exp(j2\pi a \cos \theta' / \lambda)]}{Z_T [(Z_S + Z_L)(Z_S + Z_K) - Z_1^2]} \quad (6)$$

If a plot is made of the phase and amplitude of this voltage in polar co-ordinates, the resulting locus as θ' varies will clearly be a circle with its centre displaced from the origin. The ratio of the radius of this circle to the distance of its centre from the origin is

$$r = \left| \frac{g_2Z_1}{g_1(Z_S + Z_K)} \right| \quad (7)$$

This is a useful measure of the perturbing effect of the parasitic aerial, and will be termed the interaction ratio. Note that if the aerial not in use is open-circuited Z_K is infinite, but this ratio does not vanish as g_2 also becomes infinite in this case.

(8.3) Bearing Error produced by Given Phase Errors

If the phase at each aerial is ψ_k , it has already been shown³ that the indicated bearing γ is given by

$$\gamma = \arg \left[\sum_{k=0}^{n-1} \psi_k \exp \left(\frac{2k\pi j}{n} \right) \right] \quad (8)$$

Now let $\psi_k = \beta \cos \left(\theta - \frac{2k\pi}{n} \right) + e_k$, where $\beta \cos \left(\theta - \frac{2k\pi}{n} \right)$ is the phase in the absence of error and e_k is the error in phase.

$$\begin{aligned} & \sum \beta \cos \left(\theta - \frac{2k\pi}{n} \right) \exp \left(\frac{2k\pi j}{n} \right) \\ &= \frac{\beta}{2} \sum \left[\exp j \left(\theta - \frac{2k\pi}{n} \right) + \exp -j \left(\theta - \frac{2k\pi}{n} \right) \right] \exp \frac{2k\pi j}{n} \end{aligned}$$

$$= \frac{\beta}{2} \sum_k \left[\exp j\theta + \exp -j \left(\theta - \frac{4k\pi}{n} \right) \right] \\ = \frac{1}{2} \beta n \exp j\theta, \text{ since } n \geq 3.$$

Substituting in eqn. (8) gives

$$\gamma = \arg \left[\frac{n\beta}{2} \exp j\theta + \sum e_k \exp \left(\frac{2k\pi j}{n} \right) \right] \\ = \theta + \arg \left[\frac{n\beta}{2} + \exp -j\theta \sum e_k \exp \left(\frac{2k\pi j}{n} \right) \right]$$

So the error η in indicated bearing is given by

$$\eta = \arg \left[\frac{n\beta}{2} + \sum e_k \exp j \left(\frac{2k\pi}{n} - \theta \right) \right] \quad \dots (9)$$

If $\sum_k |e_k| \ll n\beta/2$, which will certainly be true if $|e_k| \ll 1$ for all values of k ,

$$\eta = \frac{2}{n\beta} \sum e_k \sin \left(\frac{2k\pi}{n} - \theta \right) \quad \dots (10)$$

(8.4) Bearing Errors due to Aerial Interaction

(8.4.1) Effect on One Aerial of Interaction with All Other Aerials.

The treatment here resembles closely that in Section 8.2. Suppose it is desired to calculate the voltage V_k induced in the load of the k th aerial of the system when an e.m.f. V is applied to the base terminal of a distant aerial T at the bearing θ . Suppose a current I_k flows in the load of the k th aerial in this case. Then, by the principle of reciprocity, if a voltage V were applied to the base terminal of the k th aerial a current I_k would flow in the aerial T at its base terminal. If the current I_k can be calculated in this second problem, the current in the original problem is known, and hence the voltage V_k in the load of the k th aerial.

Suppose now that the k th aerial is terminated in an impedance Z_A and that all the other aerials in the ring are terminated in Z_B . Consider the inverse problem, where V is applied to the k th aerial. If the current in the $(k+j)$ th aerial is I_j , the currents are given by

$$\begin{bmatrix} Z_S + Z_A & Z_1 & Z_2 & Z_3 & \dots & Z_{n-1} \\ Z_1 & Z_S + Z_B & Z_1 & & & Z_{n-2} \\ Z_2 & & Z_1 & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_{n-1} & Z_{n-2} & & & & Z_S + Z_B \end{bmatrix} \begin{bmatrix} I_0 \\ I_1 \\ \vdots \\ \vdots \\ I_{n-1} \end{bmatrix} = \begin{bmatrix} V \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \dots (11)$$

where Z_S is the self-impedance of an aerial, and Z_j is the mutual impedance of a pair of aerials separated by an angle of $2\pi j/n$ round the ring of diameter d . Note that $Z_l = Z_{n-l}$, and $I_l = I_{n-l}$ by symmetry.

It can then be shown that

$$V_k = \sum_{l=0}^{n-1} t_l \exp \left\{ j\beta \cos \left[\theta - \frac{2(k+l)\pi}{n} \right] \right\} \quad \dots (12)$$

where $t_l = \frac{Z_A}{Z_T} g_l I_l$, and the coefficients g_l depend only on aerial current distributions and geometrical parameters.

$$g_l = g_{n-l}, \text{ hence } t_l = t_{n-l}$$

The phase, ψ_k , of this voltage is given by

$$\psi_k = \arg \left[\sum_{l=0}^{n-1} t_l \exp j \left\{ \beta \cos \left[\theta - \frac{2\pi(k+l)}{n} \right] \right\} \right] \quad \dots (13)$$

The value of the constants t_l cannot be accurately determined since the necessary calculations of mutual impedance and aerial current distributions have not been made. However, the magnitude of $|t_l/t_0|$ would be expected to be of the same order as that of the single-aerial interaction ratio defined earlier.

$$\psi_k = \arg \left[t_0 \exp \left[j\beta \cos \left(\theta - \frac{2k\pi}{n} \right) \right] \sum_{l=1}^{n-1} \frac{t_l}{t_0} \right. \\ \times \exp j\beta \left\{ \cos \left[\theta - \frac{2\pi(k+l)}{n} \right] - \cos \left(\theta - \frac{2k\pi}{n} \right) \right\} \\ = \beta \cos \left(\theta - \frac{2k\pi}{n} \right) + \phi_0 + \arg \left[1 + \sum_{l=1}^{n-1} r_l \exp (j\phi_l) \right. \\ \times \exp j\beta \left\{ \cos \left[\theta - \frac{2\pi(k+l)}{n} \right] - \cos \left(\theta - \frac{2k\pi}{n} \right) \right\} \left. \right] \quad \dots (14)$$

where $r_l \exp (j\phi_l) = t_l/t_0$, $l = 1, 2, \dots, n-1$, and $\phi_0 = \arg t_0$. r_l is taken to be real and positive. As previously noted, r_l is closely related to the interaction ratio. Note that

$$r_l = r_{n-l}, \phi_l = \phi_{n-l}$$

The constant term ϕ_0 may be neglected, since it will not affect the indicated bearing. The error in phase e_k is given by

$$e_k = \arg \left(1 + \sum_{l=1}^{n-1} r_l \exp j\theta_{kl} \right) \quad \dots (15)$$

where

$$\theta_{kl} = \beta \left\{ \cos \left[\theta - \frac{2(k+l)\pi}{n} \right] - \cos \left(\theta - \frac{2k\pi}{n} \right) \right\} + \phi_l \\ = 2\beta \sin \frac{l\pi}{n} \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] + \phi_l$$

$$\text{Now write } R \exp j\chi = \sum_{l=1}^{n-1} r_l \exp j\theta_{kl} \quad \dots (16)$$

where R is real and positive. If R is small compared with unity, as may well happen in a practical system, eqn. (15) becomes

$$e_k = R \sin \chi \\ = \sum_{l=1}^{n-1} r_l \sin \left\{ 2\beta \sin \frac{l\pi}{n} \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] + \phi_l \right\} \quad (17)$$

and the errors will be small, since R is small.

(8.4.2.) Bearing Errors due to the Phase Errors.

The bearing error, η , due to small phase errors e_k has been derived in Section 8.3 and shown to be

$$\eta = -\frac{2}{n\beta} \sum_{k=0}^{n-1} e_k \sin \left(\theta - \frac{2k\pi}{n} \right) \quad \dots (18) \\ = -\frac{2}{n\beta} \sum_{l=1}^{n-1} r_l \sum_k \sin \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] + \phi_l \right\} \\ \times \sin \left(\theta - \frac{2k\pi}{n} \right) \quad \dots (19)$$

where

$$p_l = 2\beta \sin \frac{l\pi}{n} \quad . \quad . \quad . \quad (19)$$

now in eqn. (18)

$$\begin{aligned} & \sin \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] + \phi_l \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &= \cos \phi_l \sum_{k=0}^{n-1} \sin \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &+ \sin \phi_l \sum_{k=0}^{n-1} \cos \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \quad . \quad . \quad . \quad (20) \end{aligned}$$

now

$$\begin{aligned} & \sum_{k=0}^{n-1} \sin \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &= \sum_{k=0}^{n-1} \sum_{j=0}^{\infty} 2J_{2j+1}(p_l) \sin \left\{ (2j+1) \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &= \sum_{j=0}^{\infty} J_{2j+1}(p_l) \sum_{k=0}^{n-1} \left\{ \cos \left[2j\theta - \frac{4jk\pi}{n} - \frac{(2j+1)l\pi}{n} \right] \right. \\ &\quad \left. - \cos \left[2(j+1)\theta - \frac{4(j+1)k\pi}{n} - \frac{(2j+1)l\pi}{n} \right] \right\} \quad (21) \end{aligned}$$

The first sum of cosines with respect to k must vanish unless $2j$ is integral multiple of n , i.e. unless

$$2j = 0, n, 2n, \dots$$

We will suppose now that n is even. The values

$$j = 0, \frac{n}{2}, n, \frac{3n}{2}, \dots$$

will all give a non-zero sum and correspond to the values

$$s = 0, 1, 2, 3, \dots$$

$$l = sn/2.$$

Hence

$$\begin{aligned} & \sum_{k=0}^{n-1} J_{2j+1}(p_l) \sum_{k=0}^{n-1} \cos \left[2j\theta - \frac{4jk\pi}{n} - \frac{(2j+1)l\pi}{n} \right] \\ &= \sum_{s=0}^{\infty} J_{sn+1}(p_l) n \cos \left[sn\theta - \left(s + \frac{1}{n} \right) l\pi \right] \quad . \quad (22) \end{aligned}$$

Similarly

$$\begin{aligned} & \sum_{k=0}^{n-1} J_{2j+1}(p_l) \sum_{k=0}^{n-1} \cos \left[2(j+1)\theta - \frac{4(j+1)k\pi}{n} - \frac{(2j+1)l\pi}{n} \right] \\ &= \sum_{s=1}^{\infty} J_{sn-1}(p_l) n \cos \left[sn\theta - \left(s - \frac{1}{n} \right) l\pi \right] \quad . \quad (23) \end{aligned}$$

eqn. (21) becomes

$$\begin{aligned} & \sin \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &= n \sum_{s=0}^{\infty} J_{sn+1}(p_l) \cos \left[sn\theta - \left(s + \frac{1}{n} \right) l\pi \right] \\ &\quad - n \sum_{s=1}^{\infty} J_{sn-1}(p_l) \cos \left[sn\theta - \left(s - \frac{1}{n} \right) l\pi \right] \quad . \quad (24) \end{aligned}$$

n is even.

The other sum in eqn. (20) is

$$\begin{aligned} & \sum_{k=0}^{n-1} \cos \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) \\ &= \sum_{j=1}^{\infty} J_{2j}(p_l) \sum_{k=0}^{n-1} \left\{ \cos \left[(2j-1)\theta - (2j-1)\frac{2k\pi}{n} - \frac{2jl\pi}{n} \right] \right. \\ &\quad \left. - \cos \left[(2j+1)\theta - (2j+1)\frac{2k\pi}{n} - \frac{2jl\pi}{n} \right] \right\} \quad . \quad (25) \end{aligned}$$

The first sum of cosines with respect to k will vanish unless

$$2j-1 = 0, n, 2n, \dots$$

i.e. unless $j = \frac{1}{2}, \frac{n+1}{2}, \frac{2n+1}{2}, \dots$

If n is even none of these values for j is integral; hence the sum must vanish completely. The same is true of the second sum. Hence

$$\sum_{k=0}^{n-1} \cos \left\{ p_l \sin \left[\theta - \frac{(2k+l)\pi}{n} \right] \right\} \sin \left(\theta - \frac{2k\pi}{n} \right) = 0 \quad (26)$$

provided that n is even. Hence from eqns. (18), (20), (24), (25) and (26) we have that, for n even,

$$\begin{aligned} \eta &= -\frac{2}{\beta} \sum_{l=1}^{n-1} r_l \cos \phi_l \left\{ \sum_{s=0}^{\infty} J_{sn+1}(p_l) \cos \left[sn\theta - \left(s + \frac{1}{n} \right) l\pi \right] \right. \\ &\quad \left. - \sum_{s=1}^{\infty} J_{sn-1}(p_l) \cos \left[sn\theta - \left(s - \frac{1}{n} \right) l\pi \right] \right\} \quad . \quad . \quad . \quad (27) \end{aligned}$$

Consider now

$$\begin{aligned} & \sum_{l=1}^{n-1} r_l \cos \phi_l \sum_{s=0}^{\infty} J_{sn+1}(p_l) \cos \left[sn\theta - \left(s + \frac{1}{n} \right) l\pi \right] \\ &= \sum_{s=0}^{\infty} \left[\sum_{l=1}^{n/2-1} \left\{ r_l \cos \phi_l J_{sn+1} \left(2\beta \sin \frac{l\pi}{n} \right) (-1)^{sl} \cos \left(sn\theta - \frac{l\pi}{n} \right) \right. \right. \\ &\quad \left. \left. + r_{n-l} \cos \phi_{n-l} J_{sn+1} \left(2\beta \sin \frac{l\pi}{n} \right) \right. \right. \\ &\quad \left. \left. (-1)^{s(n-l)} \cos \left[sn\theta - \frac{(n-l)\pi}{n} \right] \right\} \right. \\ &\quad \left. + r_{n/2} \cos \phi_{n/2} J_{sn+1} \left(2\beta \sin \frac{\pi}{2} \right) (-1)^{\frac{sn}{2}} \cos \left(sn\theta - \frac{\pi}{2} \right) \right] \\ &= \left\{ \sum_{s=0}^{\infty} \left[\sum_{l=1}^{n/2-1} r_l \cos \phi_l J_{sn+1} \left(2\beta \sin \frac{l\pi}{n} \right) (-1)^{sl} \right. \right. \\ &\quad \left. \left. 2 \sin (sn\theta) \sin \frac{l\pi}{n} \right] + r_{n/2} \cos \phi_{n/2} J_{sn+1} (2\beta) (-1)^{\frac{sn}{2}} \sin (sn\theta) \right\} \\ &= \sum_{s=1}^{\infty} \sin (sn\theta) \sum_{l=1}^{n/2} (2 - \delta_{l, n/2}) (-1)^{sl} r_l \cos \phi_l \\ &\quad \times J_{sn+1} \left(2\beta \sin \frac{l\pi}{n} \right) \sin \frac{l\pi}{n} \quad . \quad (28) \end{aligned}$$

since the $s = 0$ term vanishes. The Kronecker δ symbol is used here to shorten the expressions. ($\delta_{p,q} = 0$ if $p \neq q$
 $= 1$ if $p = q$)

The other sum involved in eqn. (27) can be evaluated in a similar way, namely

$$\sum_{l=1}^{n-1} r_l \cos \phi_l \sum_{s=1}^{\infty} J_{sn-1}(p_l) \cos \left[sn\theta - \left(s - \frac{1}{n} \right) l\pi \right] \\ = - \sum_{s=1}^{\infty} \sin(sn\theta) \sum_{l=1}^{n/2} (2 - \delta_{l,n/2}) (-1)^s r_l \cos \phi_l J_{sn-1} \left(2\beta \sin \frac{l\pi}{n} \right) \\ \times \sin \frac{l\pi}{n} \quad (29)$$

Hence

$$\eta = - \frac{2}{\beta} \sum_{s=1}^{\infty} \sin(sn\theta) \sum_{l=1}^{n/2} (2 - \delta_{l,n/2}) (-1)^s r_l \cos \phi_l \\ \times \sin \frac{l\pi}{n} \left[J_{sn+1} \left(2\beta \sin \frac{l\pi}{n} \right) + J_{sn-1} \left(2\beta \sin \frac{l\pi}{n} \right) \right] \\ = - \frac{2n}{\beta^2} \sum_{s=1}^{\infty} s \sin(sn\theta) \sum_{l=1}^{n/2} (2 - \delta_{l,n/2}) (-1)^s r_l \cos \phi_l \\ \times J_{sn} \left(2\beta \sin \frac{l\pi}{n} \right) \quad (30)$$

A formula when n is odd can be derived in a similar way to that when n is even.

(8.5) Calculation of the Polarization Error in the Phase due to a Feeder from the Experiments made with a Loop Transmitter

Suppose that the loop is mounted on a mast in line with two aerials A and B (see Fig. 5). A dummy feeder is directed from B so as to make an angle of χ with BA. Let the plane of the loop make an angle α from the vertical. Suppose now that the voltages at the base of aerials A and B are V_A, V_B respectively, and that the phase difference ($\arg V_B - \arg V_A$) and the amplitude ratio $|V_B|/|V_A|$ have values given in Table 5 under the stated conditions.

Table 5

PHASE DIFFERENCE AND AMPLITUDE RATIO OF THE SIGNALS PICKED UP BY THE AERIALS FOR VARIOUS VALUES OF α AND χ

α	0	α_0	—	0	α_0	$-\alpha_0$
$\arg V_B - \arg V_A$	χ_0	χ_0	χ_0	$-\chi_0$	$-\chi_0$	$-\chi_0$
$ V_B / V_A $	r_{VL}	r_{LL}	r_{RL}	r_{VR}	r_{LR}	r_{RR}

Now when $\alpha = 0, \chi = \pm \chi_0$ suppose that

$$V_A = V_{01} \exp j\phi'_1 \\ V_B = V_{02} \exp j\phi'_2$$

where $V_{01}, V_{02}, \phi'_1, \phi'_2$ are all real, and $V_{01}, V_{02} > 0$.

Then $\psi_{VL} = \phi'_2 - \phi'_1 = \psi_{VR} = \psi_V$, say . . . (31)

Also $r_{VL} = V_{02}/V_{01} = r_{VR} = r_V$, say . . . (32)

Now when $\alpha = \alpha_0, \chi = \chi_0$ suppose that

$$V_A = V_{01} \exp j\phi'_1 \cos \alpha_0 [1 + k_3 \exp j(\phi'_3 - \phi'_1)] \quad (33)$$

$$V_B = V_{02} \exp j\phi'_2 \cos \alpha_0 [1 + k_4 \exp j(\phi'_4 - \phi'_2) + k_6 \exp j(\phi'_6 - \phi'_2)] \quad (34)$$

where k_3, k_4 and k_6 are real and have the sign of $\tan \alpha_0$. Here the terms in k_3 and k_4 are due to the effect of horizontally polarized radiation on the site, excluding the dummy feeder and will change sign if the sign of α is altered. The term in k_6 is due to some form of coupling between aerial B and the dummy feeder and will change sign with an alteration of sign in either α or χ . It will be assumed that the polarization error is small i.e. that $k_3, k_4, k_6 \ll 1$.

Eight such relations can be obtained by taking all pairs of values from $\alpha = \pm \alpha_0, \chi = \pm \chi_0$, and from these the values of ψ_{LL}, r_{LL} , etc., can be found.

Now suppose for the moment that there are no errors due to the site, i.e. that $k_3 = k_4 = 0$. A method closely analogous to that used with the Adcock system can be employed to give the polarization error in the phase difference. The in-phase component of error due to the presence of the feeder may be defined as

$$e_P = \frac{\psi_{LL} - \psi_{RL}}{2}$$

In these circumstances it can be shown that

$$e_P = k_6 \sin(\phi'_6 - \phi'_2) \quad (35)$$

The total polarization error e_T is obtained by adjusting the phase of the horizontally polarized wave relative to the vertically polarized wave to maximize the in-phase error. Here this process involves varying ϕ'_6 , the phase of the induced signal in aerial B due to the horizontal electric field, so as to make $|e_P|$ maximum. Clearly this occurs when $\phi'_6 = \phi'_2 \pm \pi/2$, in which case $|e_P| = k_6 = e_T$, the total error. Now if we define "quadrature" error e_Q due to the dummy feeder by the relation

$$e_T^2 = e_P^2 + e_Q^2 \quad (36)$$

we may take the value of e_Q to be given by

$$e_Q = k_6 \cos(\phi'_6 - \phi'_2) \quad (37)$$

To return to the general case where there are errors due to the site; what is required is a method for deducing from the observable quantities given in Table 5 the total, in-phase and quadrature errors due to the presence of the dummy feeder given by eqns. (35)–(37). This can be done; it is found that the in-phase error can be calculated from

$$e_P = \frac{1}{2}(\psi_{LL} - \psi_{LR}) \quad (38)$$

or from

$$e_P = -\frac{1}{2}(\psi_{RL} - \psi_{RR}) \quad (39)$$

The formula for calculating e_Q is

$$e_Q = \frac{1}{2r_V}(r_{LL} - r_{LR}) \quad (40)$$

or

$$e_Q = -\frac{1}{2r_V}(r_{RL} - r_{RR}) \quad (41)$$

e_T can now be obtained from eqn. (36).

In practice, e_P was taken to be the mean of the values obtained from eqns. (38) and (39), and e_Q the mean of those obtained from eqns. (40) and (41).

AN AUDIO-FREQUENCY DYNAMOMETER WATTMETER

By A. H. M. ARNOLD, Ph.D., D.Eng., and J. J. HILL, B.Sc., Associate Members.

(The paper was first received 7th October, and in revised form 29th November, 1955. It was published as an INSTITUTION MONOGRAPH in March, 1956.)

SUMMARY

Negative-feedback amplifiers are used to supply the currents to the terminals of a nickel-iron-cored dynamometer, and by this means, indications proportional to current, voltage or power in an external circuit may be obtained from the dynamometer at any frequency between 50 c/s and 20 kc/s. The signal voltages required at the amplifier input terminals for full-scale instrument deflection are 2 volts. A voltage divider, having a resistance of 1000 ohms/volt and time-constant ratios less than $0.02 \mu\text{H}/\text{ohm}$, is used to give eight voltage ranges up to 100 volts. The current ranges are obtained by using four terminal resistance standards of different values and having time-constants of less than $0.02 \mu\text{H}/\text{ohm}$.

The overall instrument error at any power factor and any frequency within the working range does not exceed 0.6% of full-scale deflection. Internal switching is provided so that the voltage, current and power measurement load may be determined without disturbing the load conditions. The upper limit of frequency is 20 kc/s for power measurements, except at unity power factor, but voltage and current may be measured at frequencies up to 30 kc/s.

supplies and the current shunt, which are separate from the instrument.

(2) DYNAMOMETER MOVEMENT

The greater efficiency of the nickel-iron-cored dynamometer compared with the iron-free instrument was considered to outweigh the disadvantages associated with the use of ferromagnetic materials, and a movement was specially manufactured for the N.P.L. by a well-known manufacturer. The eddy-current losses have been reduced by the employment of 0.005 in magnetic laminations in place of the usual 0.015 in laminations. A further reduction may be possible in the future when the manufacturing difficulties associated with the use of 0.002 in laminations have been overcome.

The design of the amplifiers is easier for small output currents than for large, but errors in the dynamometer from capacitance currents may be excessive at the higher frequencies if the instrument is designed for a very small current. From a study of various instruments it was decided that the best compromise would be reached if a rated current of 30 mA were adopted for both coils. This is a convenient current for a type CV450 valve to handle. It is also probably near to the best current for the dynamometer itself, since a larger current might well involve an increased error from eddy-current losses offsetting the reduction of error from capacitance.

(3) AMPLIFIERS

The load on the amplifiers is highly inductive, and there is, moreover, the complication of mutual inductance between the dynamometer coils. This latter phenomenon, however, was not found to cause any difficulties, as would be expected from the following considerations.

In the working range of frequency two conditions of operation are of interest. In the first the e.m.f. induced in one dynamometer coil is sufficient to provide the major control of current in that coil. The most extreme example of this, of course, is when there is no input voltage for the amplifier concerned. In this case the induced e.m.f. in the coil is able to send a small current through the feedback resistance, thus providing a signal voltage on the grid of the first valve which tends to reduce the feedback current to a negligible value. In the second condition of operation there is an appreciable input voltage to the amplifier. In this case the voltage at the terminals of the coil may be increased or decreased by the induced e.m.f., but the variations which can occur are well within the power of the amplifier to handle.

Outside the working frequency range there is the possibility of a stray e.m.f. of high frequency producing a current in one coil, which in turn induces an e.m.f. in the other coil. However, such an e.m.f. can be treated in the same way as any other induced e.m.f. in the coil when considering questions of stability outside the working range.

The problem, therefore, to be considered is that of achieving stability, together with the necessary performance, when supplying a highly inductive load. The working frequency range was arbitrarily fixed at 50 c/s to 20 kc/s, and the circuit finally adopted is shown in Fig. 1.

(1) INTRODUCTION

It was shown in an earlier paper¹ that the dynamometer may be used as an ammeter up to frequencies of 20 kc/s or even higher without serious loss of accuracy, provided that the rated current was chosen so that errors from eddy-current losses and capacitance currents in the coils are both kept small. However, although the accuracy may be adequate, the high volt-ampere consumption of the instrument at the higher frequencies is a disadvantage, since it may disturb the conditions in the measured circuit.

The errors of the dynamometer as a voltmeter or a wattmeter tend to increase to an excessive extent at higher frequencies on account of the high impedance of the moving coil, which makes it difficult, except for high-voltage ranges, to maintain a current in it having the correct magnitude and phase relationship to the voltage at the voltage terminals.

These disadvantages of the dynamometer may be greatly reduced by the use of electronic amplifiers to supply the currents to the fixed and moving coils whilst requiring only small signal voltages from the measured circuit. This principle was naturally applied to the moving-coil circuit first, and various instruments using one amplifier have been described.^{2,3,4} An instrument using amplifiers for both coils has been constructed for operation at a single frequency.⁵ A new instrument has now been constructed at the National Physical Laboratory to operate over the whole range of frequency from 50 c/s to 20 kc/s, and it is thought that a description of its design would be of interest.

The accuracy achieved closely approximates to that of British Standard precision grade, and consecutive readings of voltage, current and power in a circuit may be obtained without appreciable disturbance of the measured circuit.

The instrument is semi-portable, having overall dimensions of 12 in \times 13 in \times 10 in and a weight of 51 lb. These figures do not include the dimensions and weight of the necessary power

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is an official communication from the National Physical Laboratory.

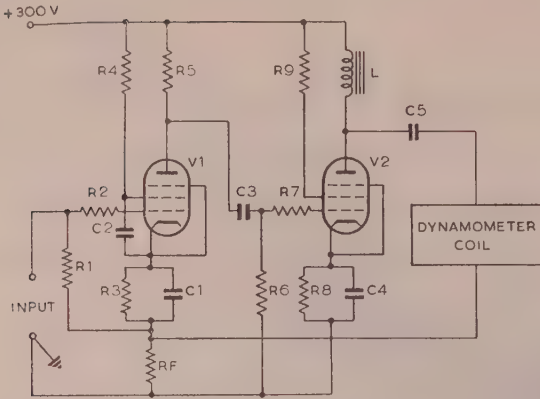


Fig. 1.—Amplifier circuit.

R1 1 MΩ, ½ W (high stability).	C1 100 μF, 25 V (electrolytic).
R2 2.2 kΩ, ½ W.	C2 0.5 μF, 250 V (paper).
R3 390 Ω, 1 W (high stability).	C3 2 μF, 350 V (paper).
R4 100 kΩ, ½ W (high stability).	C4 500 μF, 12 V (electrolytic).
R5 27 kΩ, 1 W (high stability).	C5 40 μF, 400 V (paper).
R6 120 kΩ, ½ W (high stability).	L 50 H, 80 mA.
R7 2.2 kΩ, ½ W.	V1 6X4.
R8 115 Ω, 2 W.	V2 6X5.
R9 100 Ω, ½ W (high stability).	
RF 66 Ω, Non-inductive wire wound.	

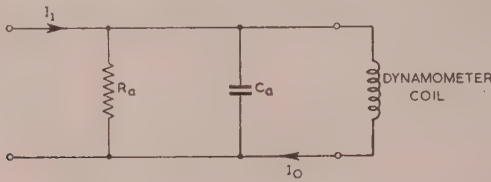


Fig. 2.—Equivalent circuit for last stage of amplifier at high frequencies.

R_a = Anode resistance.
 C_a = Total capacitance shunting valve.
 I_1 = Pentode current source.
 I_0 = Current through feedback resistance.

Considering, first, frequencies above the working range the output stage may be represented by Fig. 2. It is convenient for the moment to consider the dynamometer coil as a pure inductance. The capacitance C_a includes the valve capacitances, the capacitance to earth of the condenser C5 (see Fig. 1), the self-capacitance of the inductor L in the d.c. supply to the anode, and its capacitance to earth.

This inductance resonates with its self-capacitance within the working frequency range, so that above the working frequency it is behaving as a capacitance. The total capacitance C_a is thus appreciable, and resonance with the dynamometer-coil inductance will occur before the amplifier feedback factor has fallen to unity. As the frequency is increased above resonance the phase shift between the current, I_1 , of the last valve and the current in the feedback resistance, I_0 , will approach 180° , and since there must be some phase shift in the first stage, instability will result. This instability is avoided in practice by the effects of the resistance and self-capacitance of the dynamometer with the effect of added capacitance and resistance shunts across it. Fig. 3 shows the equivalent circuit of Fig. 2 in more detail. With this circuit the ratio of I_1/I_0 is given by eqn. (1):

$$\frac{I_1}{I_0} = \frac{\frac{1}{R_a} + \frac{1}{R_b} + \frac{R_d}{Z_d^2} + j\omega \left(C_a + C_b - \frac{L_d}{Z_d^2} \right)}{\frac{1}{R_b} + \frac{R_d}{Z_d^2} + j\omega \left(C_b - \frac{L_d}{Z_d^2} \right)} \quad (1)$$

where $Z^2 = R_d^2 + \omega^2 L_d^2$.

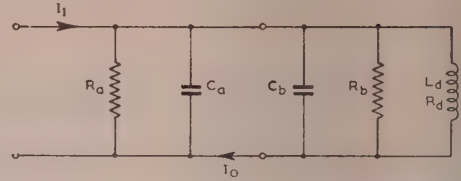


Fig. 3.—Equivalent circuit for last stage of amplifier at high frequencies showing lumped circuit components of dynamometer coil.

C_b = Self-capacitance plus added shunt capacitance.
 R_b = Added shunt resistance.
 L_d = Inductance of coil.
 R_d = Resistance of coil.

Let ϕ = Phase angle between I_1 and I_0 .

$$G_a = \frac{1}{R_a}$$

$$G = \frac{1}{R_b} + \frac{R_d}{Z_d^2}$$

Then

$$\tan \phi = \frac{\omega G C_a - \omega G_a \left(C_b - \frac{L_d}{Z_d^2} \right)}{G(G + G_a) + \omega^2 \left(C_b - \frac{L_d}{Z_d^2} \right) \left(C_a + C_b - \frac{L_d}{Z_d^2} \right)} \quad (2)$$

The first resonance occurs when $C_a + C_b = L_d/Z_d^2$. Both the numerator and denominator of eqn. (2) are positive so that the phase angle ϕ is less than 90° and no instability can occur. There may, however, be an unwelcome gain in amplification and it is desirable that this resonance should occur well outside the working frequency range. The capacitor C5 and the inductor L (see Fig. 1) were therefore separated from the chassis by ½ in-thick insulating supports in order to reduce their capacitances to the chassis to an amount which would bring the first resonance to a comparatively high frequency.

At frequencies above this first resonant frequency, but below the second resonant frequency when $C_b = L_d/Z_d^2$, the right-hand term in the denominator of the right-hand side of eqn. (2) will be negative. If, as is usually the case, $G(G + G_a)$ is small, the whole denominator will be negative for most of the frequency range between the two resonant frequencies. The numerator is positive. The phase angle therefore exceeds 90° and will reach a maximum value at some frequency intermediate between the two resonant frequencies and very approximately equal to

$$1/2\pi\sqrt{[L_d(C_b + \frac{1}{2}C_a)]}$$

The tangent of the phase angle is then approximately equal to

$$\frac{4G}{C_a}\sqrt{[L_d(C_b + \frac{1}{2}C_a)]}$$

In order to avoid instability at this point, it is important first that the frequency at which this maximum phase angle occurs should be sufficiently low for the phase shift on the first amplifier stage to be small; secondly, G should be as large as possible. It is fortunate that resistance and capacitance shunts across both dynamometer coils are desirable in order to improve the dynamometer performance within the working frequency range. The resistance shunts reduce the maximum phase shift on the last amplifier stage, while the capacitance shunt lowers the frequency at which this maximum phase shift occurs. They therefore serve the purpose of insuring the high-frequency stability of the amplifiers, in addition to improving the dynamometer performance in the working frequency range.

When the second resonant frequency is reached the phase angle once more less than 90° . As the frequency is raised further the phase shift on the last amplifier stage decreases further, but the phase shift on the first stage begins to be important and the overall amplifier phase shift increases. The total phase shift cannot reach 180° , however, so that there is no possibility of instability. The effect of mutual inductance between the dynamometer coils is to alter the frequency at which resonance occurs, but it has little effect on the maximum value of the phase shift. This conclusion is borne out in practice, since no sign of instability can be observed for any relative currents in the two coils.

Over the whole of the working frequency range, except below 10 c/s, the feedback loop gain is approximately 115 and is constant to within 5%. There is a fall in gain between 200 c/s and 50 c/s of about 10%, which affects the performance with feedback by less than 0.1%. This small error is taken account of, together with the dynamometer errors, in the overall adjustment of the whole equipment described in Section 5. There is also phase shift of about 25° at 50 c/s, which could introduce a power-factor error of up to 0.5% of full-scale deflection. Fortunately, this error is neutralized by an equal error in the other amplifier, and the power-factor error from phase shift will not exceed 0.1% provided that the performances of the two amplifiers do not differ by more than 20%.

Below the working frequency range, phase shift occurs on account of four factors—namely the feedback introduced by the series impedance of the capacitors C1 and C4 (Fig. 1), the change in the relative impedances of the dynamometer circuit including the capacitor C5 and the d.c. supply circuit including the inductor (Fig. 1), and finally the effect of the increasing impedance of the coupling capacitor C3 (Fig. 1). The phase shift due to C1 reaches a maximum value of about 40° at 4 c/s, while that due to C4 reaches a maximum value of about 30° at 3 c/s. The phase shift on account of the parallel output paths reaches a maximum value of about 110° at 2 c/s. To avoid instability it is therefore important that the phase shift due to the coupling capacitor C3 should be kept small until the frequency falls below 2 c/s. This has been achieved by the use of a large capacitor giving a phase shift at 2 c/s of 15° . It was found experimentally that there was maximum phase shift at 2 c/s of 162° , which agrees well with the calculated value of 166° .

(4) VOLTAGE DIVIDER AND CURRENT SHUNTS

The dynamometer with its two amplifiers constitutes a single-range instrument, which may be used as a voltmeter giving full-scale deflection when a voltage of 2 volts is applied to the inputs of the two amplifiers, or as an ammeter if the current is passed through a resistive shunt to give voltage drop of 2 volts and if its voltage is then applied to the amplifier inputs. The instrument cannot be used as a wattmeter on the 2-volt range, since all the applied voltage would be absorbed by the drop in the resistance shunt carrying the circuit current. It may, however, be used as a wattmeter on higher circuit voltages if a voltage divider is provided to give the necessary 2 volts for one amplifier. The use of a divider is a disadvantage in that the input impedance of the divider is very much less than that of the amplifier. Instruments designed for operation at power frequencies may be fitted with a capacitance divider having a high impedance and negligible power consumption. Such dividers are both cheap and efficient. However, the current taken by a capacitance divider increases in direct proportion to the frequency, and it is not practicable to build one having the necessary accuracy with a small current consumption at 20 kc/s. It was therefore necessary to construct a resistance divider, and in order

to obtain the requisite accuracy and stability, wire-wound units were used and not high-stability carbon units. This necessarily involved increased dimensions with a corresponding increase of capacitance. It was decided to provide voltage ranges up to 500 volts, and that it would not be practicable to have a resistance of more than 1 000 ohms/volt, making a total of half a megohm, if excessive errors from capacitive stray currents were to be avoided.

The divider consists of a 2-kilohm section which is connected across the input of one amplifier, and seven sections of resistance 3, 8, 18, 48, 123, 248 and 498 kilohms, respectively, all connected to one end of the 2-kilohm section and with their free ends connected to the inner terminals of concentric sockets mounted on the front of the instrument case. The junction of the 2-kilohm resistance and the other seven resistances is also connected to the inner terminal of a concentric socket. The outer terminals of all the concentric sockets are connected to the instrument chassis. Thus eight voltage ranges of 2, 5, 10, 20, 50, 125, 250 and 500 volts are provided and range changing is effected by moving a plug with leads connected to the circuit in which power is being measured to the appropriate socket. This is less convenient than using a range-changing switch, but it has been found easier to maintain the necessary high insulation resistance with a plug-and-socket arrangement than with a multi-contact switch. The three upper voltage ranges were chosen to be suitable for mains supply voltages of 110, 240 and 440 volts.

Each section of the divider consists of a single-layer coil wound on a mica card 2 in wide and 0.01 in thick. The edges of the card are covered with polystyrene tape, and no varnish or other impregnation is used on the wound card. Adjacent turns touch, but a gap of 0.5 mm is left between adjacent groups of 100 turns in order to reduce the self-capacitance of the coil. The wire used is of 0.001 in diameter, insulated with synthetic enamel and composed of an alloy of nickel, chromium, aluminium and copper having a resistivity $2\frac{1}{2}$ times that of manganin and a temperature coefficient of resistance of a few parts in a million per degree centigrade. The edges of the mica card need protection from the cutting action of the wire, but the protective material used must have high quality as an insulator. It was found that some experimental cards in which paper insulation and impregnating varnish were used, although quite satisfactory on direct current since the enamel on the wire provided adequate insulation, were unsatisfactory on alternating current on account of excessive self-capacitance. The poor insulation resulted in remote turns with a high voltage between them being brought into electrical proximity.

The voltage-divider ratios of the completed divider were measured, and the results showed that the time-constant differences between the input resistance and the tapping section were less than $0.05 \mu\text{H}/\text{ohm}$ for voltage ranges up to 50 volts, and less than $0.1 \mu\text{H}/\text{ohm}$ for the three higher voltage ranges. These results were very satisfactory for such high resistances, but a time-constant difference of $0.1 \mu\text{H}/\text{ohm}$ would introduce a zero-power-factor error of over 1% at 20 kc/s, and compensation was necessary in order to reduce this error. To compensate the 500-volt section would require the connection of a capacitor across the tapping section, and this would result in over-compensation of all the other sections. The following compromise procedure was adopted. Capacitance was connected in parallel with the tapping section to reduce the time-constant difference of the 125-volt range to less than $0.02 \mu\text{H}/\text{ohm}$. This resulted in slight over-compensation of the lower-voltage ranges. Capacitance was then connected in parallel with the 50-, 20-, 10- and 5-volt sections to reduce their time-constant differences to less than $0.02 \mu\text{H}/\text{ohm}$. The 250- and 500-volt sections were under-compensated, and this was rectified by adjusting the

position of a metal plate connected to the chassis so as to increase the earth capacitance of these sections. These adjustable plates were fixed parallel to, and at a distance of $\frac{1}{2}$ – $\frac{3}{4}$ in from, the section

resistance cards. The remaining cards were not less than 2 in from any earthed metal.

The capacitance of the screened lead connecting the tapping section to the amplifier input terminals was included in the adjustment.

The resistance shunts for providing a voltage proportional to the circuit current are not mounted inside the instrument case, but are separate units. The principles of their design have already been described.⁷

The dynamometer is mounted in a metal chassis together with an amplifier on each side. The voltage divider is mounted underneath the instrument together with a four-deck switch for switching the amplifiers so that the instrument reads voltage, current or power. The switching arrangements are shown diagrammatically in Fig. 4. The instrument may also be used as a 2-volt voltmeter of high impedance when switched to the ammeter position.

(5) ADJUSTMENT OF PERFORMANCE

The principal error in the completed instrument was the power-factor error of the dynamometer, and the most obvious method of adjustment of this error was to connect a resistor across the moving coil to increase the value of β_1 to that of β_2 (see Appendix 9). An inductor in series with this resistor would be of value in reducing the error at unity power factor and also in preventing over-compensation of the power-factor error at the higher frequencies. However, it was soon found that stability difficulties were met which had to be overcome by other devices. These stability difficulties were eliminated, and equally good performance was obtained by connecting resistance and capacitance across both coils as shown in Fig. 4. A little reduction of error was also achieved by adjustment of the feedback resistance. Table 1 shows the performance of the adjusted instrument, from which it may be seen that the maximum zero-power-factor error is

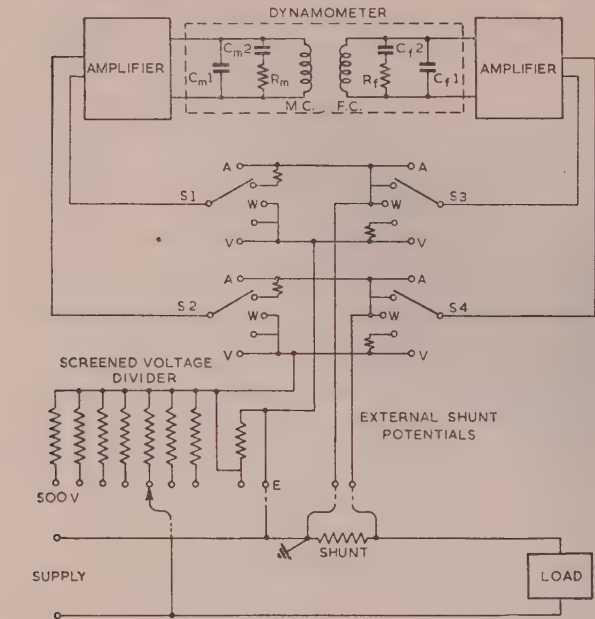


Fig. 4.—Circuit diagram for power measurements including compensation networks for dynamometer.

$R_m = 3\,600$ ohms, $C_{m1} = 24\,\mu\text{F}^*$
 $R_f = 12\,000$ ohms, $C_{m2} = 0\cdot03\,\mu\text{F}$,
 $C_{f1} = 80\,\mu\text{F}^*$,
 $C_{f2} = 0\cdot05\,\mu\text{F}$.

S1, S2, S3 and S4: Four-pole wafer switch.
* Including capacitance of connecting leads.

Table 1

OVERALL ERRORS OF COMPENSATED DYNAMOMETER WITH AMPLIFIERS AND VOLTAGE DIVIDER

Wattmeter scale reading	Circuit power factor	Error, % full-scale value						
		50 c/s	300 c/s	1 kc/s	4 kc/s	12 kc/s	16 kc/s	20 kc/s
(a) Instrument tested as a wattmeter at rated voltage on any range.								
100	Unity	-0.1	-0.1	0.0	+0.1	-0.1	-0.2	-0.2
70	Unity	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1
50	Unity	+0.2	+0.1	+0.1	0.0	-0.1	0.0	0.0
30	Unity	+0.2	+0.2	+0.2	0.0	0.0	0.0	0.0
50	0.5 (lead)	+0.4	+0.3	+0.4	+0.4	+0.2	+0.2	+0.2
50	0.5 (lag)	0.0	-0.1	-0.2	-0.4	-0.3	-0.1	+0.2
0	Zero (lead)	+0.5	+0.5	+0.3	0.0	-0.1	-0.1	0.0
0	Zero (lag)	0.0	0.0	-0.2	-0.6	-0.4	+0.1	+0.6
Voltmeter scale reading		Error, % indicated value						
		50 c/s	300 c/s	1 kc/s	4 kc/s	12 kc/s	20 kc/s	30 kc/s
(b) Instrument tested as a voltmeter or ammeter on any range.								
100		-0.2	-0.2	-0.1	0.0	-0.1	-0.2	0.0
80		0.0	0.0	0.0	-0.1	-0.1	+0.1	+0.6
60		+0.1	0.0	0.0	-0.2	-0.1	+0.1	+0.6
40		0.0	-0.2	-0.2	-0.4	-0.4	-0.2	+0.4

Two scales are marked on the scale plate, one divided approximately uniformly and one divided approximately according to a square law.

6% of full-scale deflection and the maximum unity-power-factor error is 0.2% of full-scale deflection.

(6) DEFECTS OF THE COMPLETE INSTRUMENT

The instrument described is capable of a reasonable accuracy over a wide frequency range, but it does not possess all the desirable qualities. Some of the defects may be eliminated by reducing the range of frequency operation and some must await further development. The principal defects are briefly considered below.

(6.1) Accuracy

The power-factor error of the uncompensated dynamometer might be reduced by the employment of thinner nickel-iron laminations. This would reduce the amount of compensation necessary. The power-factor error arising from the change of self-inductance of the moving coil with deflection might be reduced by a reduction of the moving-coil ampere-turns. Alternatively, it might be possible to introduce a compensating device such as a magnetic shunt, in order to eliminate this effect.

(6.2) Earthing

It is necessary to earth the chassis in order to avoid capacitance effects from the observer's body when reading the instrument, and this means that one lead in the circuit in which power is being measured must also be earthed. This would be a disadvantage in some cases. When the instrument is only required for a restricted range of frequency a solution may be found in the use of transformers to isolate the measured circuit from the instrument.

(6.3) Voltage Drop in the Current Shunt

The drop of 2 volts in the current shunt might be excessive in some cases. This could be reduced to $1\frac{1}{2}$ volts or even 1 volt by alteration of the feedback resistance, with perhaps a small restriction of the working frequency range. If a further reduction were required an extra stage of amplification would be necessary. Alternatively, the use of a current transformer with the shunt connected in the secondary would provide an acceptable solution in many cases.

(6.4) Impedance of the Voltage Divider

The impedance of the voltage divider might be increased by the use of high-stability carbon resistors, but it is doubtful whether these are yet sufficiently reliable for high-class work. The alternative of range changing by alteration of the feedback resistance might be used for low-voltage ranges, but there are considerable difficulties in the way of applying it to the higher-voltage ranges.

(6.5) Limitation of Frequency Range

The upward extension of the frequency range might be possible at the cost of reducing the instrument torque. The use of a suspended movement in place of a pivoted movement is an obvious development. There would be considerable difficulty in reducing the time-constants of the voltage divider below their present values, and some loss of accuracy at higher frequencies would almost certainly have to be accepted.

(7) ACKNOWLEDGMENTS

The work described has been carried out as part of the research programme of the National Physical Laboratory, and the paper published by permission of the Director of the Laboratory. The authors desire to acknowledge the assistance rendered in the experimental observations by Mr. A. P. Miller.

(8) REFERENCES

- (1) ARNOLD, A. H. M.: "Audio-Frequency Power Measurements by Dynamometer Wattmeters," *Proceedings I.E.E.*, Paper No. 1653 M, April, 1954 (102, B, p. 192).
- (2) MACFADYEN, K. A., and HILL, N. D.: "A Precision Thermionic A.C. Test Set," *G.E.C. Journal*, 1941, 11, p. 182.
- (3) ARNOLD, A. H. M.: "Performance Limits in Electrical Instruments," *Proceedings I.E.E.*, Paper No. 1128 M, March, 1951 (98, Part II, p. 701).
- (4) EDMUNDSON, D.: "Electrical and Magnetic Measurements in an Electrical Engineering Factory," *ibid.*, Paper No. 1703 M, September, 1954 (102 B, p. 427).
- (5) BOAST, W. B.: "An Electronic Wattmeter," *Proceedings of the National Electronics Conference* (Chicago), 1952, 8, p. 716.
- (6) BUTTERWORTH, S.: "Capacity and Eddy Current Effects in Inductometers," *Proceedings of the Physical Society*, 1921, 33, p. 312.
- (7) ARNOLD, A. H. M.: "Alternating-Current Resistance Standards," *Proceedings I.E.E.*, Paper No. 1512 M, June, 1953 (100, Part II, p. 319).

(9) APPENDIX

The simple theory of the dynamometer is concerned with demonstrating that the response of the instrument is approximately proportional to power, and with deriving an expression for the power-factor error arising from the inductance in the voltage circuit. This power-factor error is practically eliminated by using an amplifier to force the correct current through the moving coil irrespective of the change in its impedance with frequency. By this means the useful frequency range of the instrument is greatly extended, and it is desirable to consider what other errors limit the accuracy and frequency range. It will be assumed in the first place that it is the function of the amplifiers to deliver the correct currents, I_m and I_v , to the terminals of the dynamometer coils, and the function of the dynamometer to make the correct response. In practice it may be possible, to a limited extent, to balance dynamometer errors against amplifier errors, but this need not be considered theoretically.

The torque on the moving coil is approximately proportional to the scalar product of the currents in the fixed and moving coils multiplied by the rate of change of the mutual inductance between the coils with respect to angular deflection of the moving coil ($dM/d\theta$). This latter quantity, $dM/d\theta$, is usually designed to be nearly independent of θ , so that the torque may be assumed to have the same coefficient of proportionality to the scalar product of the coil currents for all values of θ . Now, the effective coil currents producing torque will differ from the currents at the coil terminals on account of the effect of the coil capacitance and its associated dielectric loss. This may be allowed for, and since capacitance is only of importance at the higher frequencies when the inductive reactance of the coils is large compared with their resistance, the capacitance may be represented in an equivalent circuit diagram by a lumped capacitance connected across a pure inductance. The dielectric loss may usually be neglected.

In addition to the coil self-capacitances there is capacitance between coils and between each coil and the screen. Butterworth⁶ has shown that the effect of these capacitances may be represented by lumped capacitances connected between the coil terminals and between the coil terminals and the screen. In this case the screen and one terminal of each coil are maintained near to the same potential by the connections to the amplifiers, so that it is possible to reduce the system of capacitances to

three—an effective self-capacitance of each coil and a capacitance between the terminal of each coil which is not near screen potential. Each of these capacitances includes the effect of capacitance between the appropriate points via the screen. The three capacitances are shown in the equivalent circuit diagram (Fig. 5). Two coil terminals are shown connected together, although there is actually a small voltage between them.

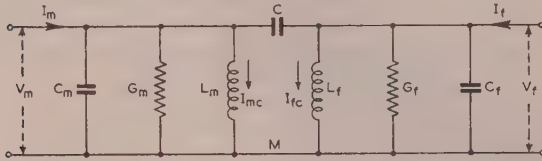


Fig. 5.—Equivalent circuit diagram of dynamometer.

Hysteresis loss in the iron produces a constant phase shift between the coil current and the magnetic field produced by it. Assuming that the hysteresis loss is proportional to the square of the flux density, it may be represented in the equivalent circuit diagram by a conductance connected in parallel with the fixed coil inductance. This conductance is inversely proportional to the frequency. No similar conductance is required for the moving coil since the hysteresis losses produced by the current in this coil do not affect the torque.

The textbook treatment of eddy-current losses is to assume an eddy-current path of resistance R_e , inductance L_e and impedance $Z_e = R_e^2 + \omega^2 L_e^2$ having mutual inductance M_e with the fixed coil. Then the eddy current is given by

$$I_e = -M_e I_f \omega (\omega L_e + jR_e) / Z_e^2 \quad (1)$$

The torque-producing effect of I_e is a fraction, say K , of that of an equal current in the fixed coil, so that the total torque is given by

$$T = (I_f + KI_e) \cdot I_m \quad (2)$$

On this basis the eddy-current effect may be allowed for in the equivalent circuit diagram by a conductance and a negative capacitance connected in parallel across the fixed coil inductance having the values given by

$$\text{Effective conductance, } G_e = KM_e R_e / Z_e^2 L_f \quad (3)$$

$$\text{Effective capacitance, } C_e = -KM_e L_e / Z_e^2 L_f \quad (4)$$

This approximation is often satisfactory when only a small frequency range is under investigation or when only a qualitative treatment is required. It is less satisfactory for a large frequency range, especially with instruments having ferromagnetic cores. The causes for this are probably, first, that there are several eddy-current paths having different ratios of R_e/L_e , and secondly, that the resistance R_e , and to a lesser extent the inductances M_e and L_e , vary with frequency. Eqns. (3) and (4) should therefore be replaced by eqns. (5) and (6), where R_{en} , M_{en} and L_{en} are variable with frequency:

$$G_e = \sum_{n=1}^{\infty} \frac{K_n M_{en} R_{en}}{(R_{en}^2 + \omega^2 L_{en}^2) L_f} \quad (5)$$

$$C_e = - \sum_{n=1}^{\infty} \frac{K_n M_{en} L_{en}}{(R_{en}^2 + \omega^2 L_{en}^2) L_f} \quad (6)$$

Eqns. (5) and (6) provide adequate scope to account for the performance of most instruments, but are not very practical. If the frequency range is not too large and the eddy-current

torque is small, the eddy-current paths may be divided into two groups—those where the reactance is negligible compared to the resistance and those where the resistance is small, but not negligible, compared to the reactance. For the former group the effective capacitance may be neglected, while for the latter group the effective conductance may be neglected. The composite effect of the former group may then be represented in the equivalent circuit diagram by a conductance connected across the fixed coil inductance. This conductance is independent of frequency at low frequencies, but is inversely proportional to the square root of the frequency at high frequencies in accordance with the skin-effect law. The working range of frequency may usually be covered by the approximation of eqn. (7), where a and b are independent of frequency:

$$G_e = \frac{1}{a + b\sqrt{\omega}} \quad (7)$$

The latter group of eddy-current paths may be represented in the equivalent circuit diagram by a negative capacitance, given approximately by eqn. (8), where p is a constant having the dimensions L/R^2 and q is a time-constant:

$$C_e = - \frac{p}{1 + \omega^2 q^2} \quad (8)$$

The values of a , b , p and q may be deduced from the performance of the instrument or estimated from design data.

The conductance G_f , shown in Fig. 5, includes the effect of the dielectric loss, which is usually negligible, the hysteresis loss and the eddy-current loss in accordance with eqn. (7). The capacitance C_f , shown in Fig. 5, includes the effect of self-capacitance and the negative-capacitance effect of the eddy-current losses.

The currents supplied by the amplifiers are shown in Fig. 5 as I_m and I_f , while the currents which are effective in producing torque are those shown in the Figure as I_{mc} and I_{fc} . The other parameters are defined by the following equations:

$$C_m = C_{mc} + C_{ml} + C_{ma} \quad (9)$$

where C_{mc} = Moving-coil self-capacitance.

C_{ml} = Capacitance of leads from moving coil to amplifier.

C_{ma} = Any added capacitance across coil terminals.

$$G_m = G_{mc} + G_{ma} \quad (10)$$

where G_{mc} = Effective conductance across coil terminals due to dielectric loss.

G_{ma} = Any added conductance across coil terminals.

When G_{ma} is zero, G_m may usually be assumed to be zero also since G_{mc} is small.

L_m = Self-inductance of moving coil.

M = Mutual inductance between coils and is assumed to be positive when the moving coil is in the position corresponding to full-scale deflection.

L_f = Self-inductance of fixed coil.

$$G_f = G_{fc} + G_{fa} + \frac{G_h}{\omega} + \frac{1}{a + b\sqrt{\omega}} \quad (11)$$

where G_{fc} = Effective conductance across coil terminals due to dielectric loss.

G_{fa} = Any added conductance across coil terminals.

$\frac{G_h}{\omega}$ = Conductance across coil terminals to account for the effect of hysteresis.

Table 2

DYNAMOMETER CONSTANTS

Constant	Value	Remarks
L_m	3.3 mH	Determined by resonating moving coil with added capacitance at various frequencies, with coil in position of zero mutual inductance with fixed coil
L_f	8.4 mH	Determined similarly to above
M	3.4 mH (maximum)	Determined by resonating coils connected in series when in position of maximum negative mutual inductance, and also by direct measurement
G_{me}	Neglected	Not measured
G_{fc}	Neglected	Not measured
G_h	0.25 mho $\times \omega$	Determined from power-factor error of dynamometer at frequencies of 50 and 100 c/s
a	14000 ohms	Determined from power-factor error of dynamometer at frequencies from 300 c/s to 20000 c/s
b	82 ohms/ $\sqrt{\omega}$	
p	340 μ F	Determined from performance of dynamometer as an ammeter at frequencies up to 20000 c/s
q	27 μ s	
C	20 μ F	The value given is half that directly measured between the two coils, including capacitance between coils via the screen, with each coil short-circuited
C_{mc}	35 μ F	Determined from performance of dynamometer with current supplied to fixed coil only, and also deduced from resonance measurements above
C_{fc}	25 μ F	Determined similarly to above
C_{ml}	25 μ F	Direct measurement
C_{fl}	25 μ F	Direct measurement
For $f = 20$ kc/s and $M = +3.4$ mH		
$\alpha_1 = 0.0031$; $\alpha_2 = 0.0082$;		
$\alpha_3 = -0.0016$; $\alpha_4 = -0.0027$		
$\beta_1 = 0$; $\beta_2 = 0.027$; $\beta_3 = 0$;		
$\beta_4 = 0.011$		
For $f = 20$ kc/s and $M = -3.4$ mH		
$\alpha_1 = 0.0052$; $\alpha_2 = 0.0104$;		
$\alpha_3 = 0.0070$; $\alpha_4 = 0.0048$		
$\beta_1 = 0$; $\beta_2 = 0.027$;		
$\beta_3 = 0$; $\beta_4 = -0.011$		

$\frac{1}{1 + b\sqrt{\omega}}$ = Conductance across coil terminals to account for the effect of eddy currents.

$$C_f = C_{fc} + C_{fl} + C_{fa} - \frac{p}{1 + \omega^2 q^2} \quad (12)$$

where C_{fc} = Fixed coil self-capacitance.
 C_{fl} = Capacitance of leads from fixed coil to amplifier.
 C_{fa} = Any added capacitance across coil terminals.

$\frac{p}{1 + \omega^2 q^2}$ = Equivalent negative-capacitance effect of eddy currents.

C = Effective capacitance between coil terminals which are not near screen potential.

The effective currents I_{mc} and I_{fc} are related to the currents applied by the amplifiers, I_m and I_f , by eqns. (13), (14) and (15):

$$DI_{mc} = I_m(1 - \alpha_2 + j\beta_2) - I_f(\alpha_3 + j\beta_3) \quad (13)$$

$$DI_{fc} = I_f(1 - \alpha_1 + j\beta_1) - I_m(\alpha_4 + j\beta_4) \quad (14)$$

$$D = (1 - \alpha_1 - \alpha_2 + \alpha_1\alpha_2 - \alpha_3\alpha_4 - \beta_1\beta_2 + \beta_3\beta_4) + j(\beta_1 + \beta_2 - \alpha_1\beta_2 - \beta_1\alpha_2 - \alpha_3\beta_4 - \alpha_4\beta_3) \quad (15)$$

$$\begin{aligned} \alpha_1 &= \omega^2 C(L_m - M) + \omega^2 C_m L_m \\ \alpha_2 &= \omega^2 C(L_f - M) + \omega^2 C_f L_f \\ \alpha_3 &= \omega^2 C(L_f - M) - \omega^2 C_m M \\ \alpha_4 &= \omega^2 C(L_m - M) - \omega^2 C_f M \\ \beta_1 &= \omega G_m L_m \\ \beta_2 &= \omega G_f L_f \\ \beta_3 &= \omega G_m M \\ \beta_4 &= \omega G_f M \end{aligned}$$

eqns. (16) and (17) follow from eqns. (13), (14) and (15):

$$\begin{aligned} |D|^2 &= (1 - 2\alpha_1 + \alpha_1^2 + \beta_1^2)(1 - 2\alpha_2 + \alpha_2^2 + \beta_2^2) \\ &+ (\alpha_3^2 + \beta_3^2)(\alpha_4^2 + \beta_4^2) \\ &- 2(\alpha_3 - \alpha_1\alpha_3 + \beta_1\beta_3)(\alpha_4 - \alpha_2\alpha_4 + \beta_2\beta_4) \\ &+ 2(\beta_3 - \alpha_1\beta_3 - \alpha_3\beta_1)(\beta_4 - \alpha_2\beta_4 - \alpha_4\beta_2) \quad (16) \end{aligned}$$

$$\begin{aligned} |D|^2 I_{mc} \cdot I_{fc} &= I_m \cdot I_f (1 - \alpha_1 - \alpha_2 + \alpha_1\alpha_2 + \alpha_3\alpha_4 + \beta_1\beta_2 + \beta_3\beta_4) \\ &+ |I_m| |I_f| (\beta_2 - \beta_1 - \alpha_1\beta_2 + \alpha_2\beta_1 + \alpha_3\beta_4 - \alpha_4\beta_3) \sin \phi \\ &- |I_m|^2 (\alpha_4 - \alpha_2\alpha_4 + \beta_2\beta_4) - |I_f|^2 (\alpha_3 - \alpha_1\alpha_3 + \beta_1\beta_3) \quad (17) \end{aligned}$$

where ϕ is the phase angle between I_m and I_f .

Eqns. (16) and (17) show that the instrument response is a function of power only when the α and β terms are negligibly small.

At very high frequencies the α terms will be large compared with the β terms, so that the latter may be neglected and eqn. (15) simplifies to eqn. (18):

$$D \simeq (1 - \alpha_1)(1 - \alpha_2) - \alpha_3\alpha_4 \quad (18)$$

The system is in resonance when $(1 - \alpha_1)(1 - \alpha_2) = \alpha_3\alpha_4$, and the torque then rises to a high value, which is limited in practice by the finite values of the β terms. It is clearly impracticable to use the instrument near the resonant frequencies, and it is therefore desirable to estimate these. The constants of the dynamometer were measured and the values obtained are given in Table 2. The resonant frequency is dependent on the mutual inductance and varies from 170 to 210 kc/s. There is another resonant frequency band from 410 to 520 kc/s. The lowest resonant frequency, 170 kc/s, which was verified by direct measurement, is more than eight times the upper frequency limit of the instrument when used as a wattmeter and more than five

times its upper limit when used as a voltmeter or ammeter. It is interesting to consider the behaviour of the instrument when measuring a constant power, as the lower resonant-frequency band is approached from a lower frequency. Neglecting the β terms, eqn. (17) becomes eqn. (19):

$$\begin{aligned} |D|^2 I_{mc} \cdot I_{fc} &\simeq I_m \cdot I_f (|D| + 2\alpha_3\alpha_4) \\ &- \alpha_4 |I_m|^2 (1 - \alpha_2) - \alpha_3 |I_f|^2 (1 - \alpha_1) \quad (19) \end{aligned}$$

D will be a positive small quantity approaching zero. α_1 and α_2 are positive quantities less than unity. α_3 and α_4 are also less than unity but may be positive or negative according to the value of the mutual inductance M .

For a scale range from somewhat above half scale to full scale both α_3 and α_4 are negative. In this range the torque will be positive, and as it increases the pointer will increase its deflection until it reaches the stop above full-scale deflection.

For approximately half the scale range from zero deflection to a little above half scale both α_3 and α_4 are positive. The torque due to the first term on the right-hand side of eqn. (19) will be positive and opposed to the torque due to the remaining terms. The negative torque will usually predominate and the pointer will deflect below zero until it reaches the stop. There will be a small range of the instrument scale near half scale where

the positive and negative torques may be sufficiently closely balanced for the pointer to stay on the scale, but the equilibrium achieved will tend to be unstable. The movements of the pointer will be greatly reduced in practice, partly on account of the effect of the β terms and partly because, as the pointer moves, the resonant frequency changes also.

If the upper resonant frequency is approached the pointer tends to deflect to the position of zero mutual inductance for any load conditions.

Now, reverting to the working frequency range of the instrument, it may be seen from Table 2 that both the α and the β terms are small quantities compared with unity. To the first order of small quantities, eqns. (16) and (17) may be simplified to

$$|D|^2 = 1 - 2\alpha_1 - 2\alpha_2 \quad \dots \quad (20)$$

$$|D|^2 I_{mc} \cdot I_{fc} = (1 - \alpha_1 - \alpha_2) I_m \cdot I_f \\ + (\beta_2 - \beta_1) |I_m| |I_f| \sin \phi - \alpha_4 |I_m|^2 - \alpha_3 |I_f|^2 \quad (21)$$

whence

$$I_{mc} \cdot I_{fc} = (1 + \alpha_1 + \alpha_2) I_m \cdot I_f \\ + (\beta_2 - \beta_1) |I_m| |I_f| \sin \phi - \alpha_4 |I_m|^2 - \alpha_3 |I_f|^2 \quad (22)$$

When the instrument is used as an ammeter or voltmeter, $I_f = I_m$ and $\sin \phi = 0$. Eqn. (22) then simplifies to

$$I_{mc} \cdot I_{fc} = (1 + \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4) |I_m|^2 \\ = \{1 + \omega^2 [C_m(L_m + M) + C_f(L_f + M)]\} |I_m|^2 \quad (23)$$

Eqns. (22) and (23) provide a means of estimating the dynamometer response for any values of the currents at the coil terminals in terms of six instrument parameters, $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1$ and β_2 . The validity of these equations depends on the assumption that the only significant torque on the movement arises from the change of mutual inductance with deflection. This, unfortunately, is not quite true. There is a small change in the self-inductance of the moving coil with deflection, which gives rise to a small unwanted torque.

Measurements indicate that the self-inductance is a maximum when the coil is in the position of zero mutual inductance with the fixed coil. Eqn. (24) may be adopted to represent the self-inductance of the coil with sufficient accuracy for determining the torque:

$$L_m = L_{m0} \left[1 - \gamma \left(\frac{M}{L_{m0}} \right)^2 \right] \quad \dots \quad (24)$$

Table 3

ERRORS OF UNCOMPENSATED DYNAMOMETER

Frequency	I_m	I_f	Error in instrument deflection		Frequency	I_m	I_f	Error in instrument deflection	
			Full-scale deflection					Full-scale deflection	
			Measured parts in 3000	Calculated				Measured parts in 3000	Calculated
kc/s	mA	mA			kc/s	mA	mA		
1	30	30	-1	0	1	30	+j30	+24	+23
	25.5	25.5	-2	0		30	-j30	-6	-5
	21	21	-1	0		2	30	+j30	+30
2	30	30	-2	-1	30		-j30	-12	-11
	25.5	25.5	-2	-1	4		30	+j30	+33
	21	21	-1	0		30	-j30	-21	-20
4	30	30	-3	-3		8	30	+j30	+48
	25.5	25.5	-3	-2	30		-j30	-39	-36
	21	21	-2	-1	10		30	+j30	+56
8	30	30	-1	-3		30	-j30	-46	-45
	25.5	25.5	-3	-2		12	30	+j30	+57
	21	21	-2	-1	30		-j30	-54	-53
12	30	30	+5	+4	16		30	+j30	+66
	25.5	25.5	+2	+2		30	-j30	-69	-70
	21	21	-1	+1		20	30	+j30	+72
16	30	30	+16	+16	30		-j30	-93	-89
	25.5	25.5	+7	+8	20		30	0	+6
	21	21	+2	+4		30	0	+6	+4
20	30	30	+31	+32		30	0	+7	+8
	25.5	25.5	+13	+18	30	0	+8	+9	
	21	21	+6	+9	30	0	+9	+10	
0.05	30	+j30	+14	+16	8	30	0	+9	+10
	30	-j30	+1	+2	4	30	0	+9	+10
	0.1	30	+j30	+15	+16	2	30	0	+9
30		-j30	+1	+2	1	30	0	+9	+9
0.3		30	+j30	+16	+18	20	0	30	-21
	30	-j30	-3	0	16	0	30	-11	-13
	0.5	30	+j30	+21	+20	12	0	30	-6
30		-j30	-2	-1	10	0	30	-3	-5
30		0	0	0	8	0	30	-2	-3
					4	0	30	0	-1
					2	0	30	0	0
					1	0	30	0	0

where γ is a constant coefficient and L_{m0} is the inductance of the moving coil in the position of zero mutual inductance.

The torque, which is proportional to the change of stored energy with deflection, may be represented as follows:

$$\frac{d}{d\theta}(\frac{1}{2}L_m|I_m|^2) = -\gamma\frac{M}{L_{m0}}|I_m|^2\frac{dM}{d\theta} \simeq -\gamma\frac{M}{L_m}|I_m|^2\frac{dM}{d\theta} \quad (25)$$

Since L_m is approximately equal to L_{m0} . This torque may be added to that given by eqn. (22) to give eqn. (26) for the total instrument torque:

$$\begin{aligned} \text{Total instrument torque} = & \frac{dM}{d\theta} \left[(1 + \alpha_1 + \alpha_2)I_m \cdot I_f \right. \\ & \left. + (\beta_2 - \beta_1)|I_m||I_f| \sin \phi - \alpha_3|I_f|^2 - \left(\alpha_4 + \gamma\frac{M}{L_m} \right) |I_m|^2 \right] \quad (26) \end{aligned}$$

For the particular dynamometer used in the N.P.L. instrument, γ had a value of 0.003.

The scale is marked so that it is free of error when the instrument is used as an ammeter at low frequencies, which means that it is based on the total instrument torque being equal to

$$\left(1 - \gamma\frac{M}{L_m}\right)I_m \cdot I_f \frac{dM}{d\theta}$$

Eqn. (27) may therefore be written for the instrument error under other conditions expressed as a fraction of full-scale deflection:

$$\begin{aligned} \frac{\text{Error in instrument deflection}}{\text{Full-scale deflection}} = & \left[\left(\alpha_1 + \alpha_2 + \gamma\frac{M}{L_m} \right) I_m \cdot I_f \right. \\ & \left. + (\beta_2 - \beta_1)|I_m||I_f| \sin \phi - \alpha_3|I_f|^2 - \left(\alpha_4 + \gamma\frac{M}{L_m} \right) |I_m|^2 \right] / I^2 \quad (27) \end{aligned}$$

where I is the current in the fixed and moving coils to give full-scale deflection at unity power factor.

Table 3 shows the measured performance of the dynamometer together with the calculated performance using appropriate values of the parameters. The reasonable agreement shown indicates that eqn. (27) is satisfactory for estimating the instrument error over a considerable range of conditions.

The dynamometer errors, although small, are not negligible, and in the complete instrument they have been reduced by compensation. The method of compensation is described in Section 5.

It may be seen from eqn. (27) that the error caused by the γ terms is only significant when $I_m \cdot I_f$ differs appreciably from $|I_m|^2$.

Two cases are of interest. The first is at low power factors when $I_m \cdot I_f$ is negligible compared to $|I_m|^2$. In such cases there is only one significant γ term in eqn. (27), and this may be determined empirically by observing the instrument deflection with current in the moving coil only. Alternatively, the instrument zero may be set with current in the moving coil only, i.e. electrical zero, when the γ term will be eliminated from the instrument error. The second case of interest is for readings above one-third of full scale. In all such readings either the mutual inductance M is small, or the difference between $I_m \cdot I_f$ and $|I_m|^2$ is small so that the γ terms do not introduce significant error. This was verified experimentally by determining the difference in the instrument error at unity power factor when first the fixed-coil current was varied while the moving-coil current was maintained constant at rated value, and when, secondly, the moving-coil current was varied while the fixed-coil current was maintained constant at

Table 4

DIFFERENCE OF DYNAMOMETER ERRORS AT UNITY POWER FACTOR

Scale reading = S	Difference of errors. Test A minus Test B	Calculated difference of error
	% of full-scale deflection	% of full-scale deflection
100	—	—
90	-0.02	-0.05
80	-0.06	-0.07
70	-0.05	-0.06
60	-0.05	-0.04
50	-0.08	0.00
40	-0.08	+0.05
30	+0.01	+0.11
20	+0.17	+0.18
10	+0.28	+0.25
0	+0.30	+0.31

Test A—Moving-coil current = 30 mA.
Test B—Fixed-coil current = 30 mA.

Calculated difference of error is

$$\begin{aligned} & \left[\left(\frac{S}{100} \right)^2 - 1 \right] \gamma \frac{M}{L_m} \times 100\% \text{ of full-scale deflection.} \\ & = 0.0062(S - 50) \left[\left(\frac{S}{100} \right)^2 - 1 \right] \% \text{ of full-scale deflection.} \end{aligned}$$

the rated value. Table 4 shows the measured and calculated results. The instrument error under the two test conditions differs from the mean by less than 0.1% at all instrument deflections from 30% upwards. The methods of measurement have been described in earlier papers.^{1,3}

RESULTANT REACTIVE POWER OF OVERHEAD LINES

By A. F. COVENTRY, B.Sc., Associate Member.

(The paper was first received 27th June, and in revised form 30th November. It was published as an INSTITUTION MONOGRAPH in April, 1956.)

SUMMARY

A method is demonstrated for arriving at simple formulae by which the resultant reactive power of overhead lines under any load condition can be quickly calculated, and in which the only line constant necessary to be known is the inductance. Tables of practical overhead-line inductances are given. The method is extended to lines containing series capacitors and also to low-voltage lines. The accuracy of the method, applied both to electrically short and long lines, is discussed fully in the Appendix.

(1) INTRODUCTION

An overhead line will act as a capacitor or a reactor depending on the load current flowing in it. Another way of stating this is to say that an overhead line will act as a reactive-power source or load, depending on the load current flowing in it. When no load current flows, all overhead lines are reactive-power sources, but as the load current increases, the reactive-power output falls, passes through a zero point and changes to a reactive load. At the zero point the line is transmitting its natural or surge-impedance load, if the power factor of the load is unity.

Let us call capacitive power Q_C , and inductive power Q_L . The output of Q_C varies as the square of the line voltage. The consumption of Q_L varies as the square of the line current. They can be subtracted arithmetically to give the resultant reactive-power performance of the line. It follows from this that the higher the voltage of transmission, the more likely is an overhead line to be a reactive-power source rather than a reactive load, because line currents tend to be of the same order whatever the voltage, and the line reactance also is independent of the voltage, except to the extent that the voltage influences spacing of conductors.

It is useful to be able to evaluate the reactive-power performance of overhead lines in the manner indicated above because it facilitates the planning of other reactive-power sources, such as synchronous condensers and capacitors, and the disposition of them in such a way as to increase the system efficiency to a desirable level. A method is demonstrated below. It is based on the assumption that line current will not differ greatly from load current and that if the receiving-end voltage is fixed, the sending-end voltage will vary, with varying load, only within the limits which are normally regarded as practical. These assumptions are valid for all "short" lines, by which is meant lines not exceeding about 100 miles in length. They are also valid for lines which are electrically long, say 100 to 300 miles long, provided that the voltage regulation of these lines is kept within acceptable limits by means of equipment such as synchronous condensers. They are not valid for electrically long lines where the effect of the inductance and capacitance of the line on load current and voltage is not mitigated in any way, and in such cases the method will not give accurate results; but in practice this is not of importance because, for obvious reasons, the voltage has to be regulated. The accuracy of the method when used for both short and long lines is discussed in Appendix 6.

When the method is used for long lines associated with synchronous condensers, it is necessary to remember that it gives the reactive-power performance of the line alone and not that of the line and synchronous condensers combined. This point is also discussed in Appendix 6.

(2) METHOD

$$Q_C \text{ (3-phase)} = 2\pi f C' l V^2$$

$$Q_L \text{ (3-phase)} = 3 \times 2\pi f L' l I^2$$

where f = Frequency.

l = Length of line, miles.

C' = Capacitance per conductor per mile, farads.

L' = Inductance per conductor per mile, henrys.

V = Line voltage, volts.

I = Line current, amp.

For overhead lines the product $L'C'$ is practically a constant and equal to 3×10^{-11} .

Therefore

$$C' = \frac{3 \times 10^{-11}}{L'} \quad \dots \quad (1)$$

Now the resultant line reactive power (in vars) is

$$Q_L - Q_C = 3 \times 2\pi f L' l I^2 - 2\pi f C' l V^2$$

Substituting for C' from eqn. (1) and changing vars to kilovars and volts to kilovolts, this becomes

$$Q_L - Q_C = 0.01884 f l \left(L' I^2 - \frac{(kV)^2}{L' \times 10^5} \right) \quad \dots \quad (2)$$

and if $f = 50$, the resultant line reactive power (in kilovars) is

$$0.942 l \left(L' I^2 - \frac{(kV)^2}{L' \times 10^5} \right)$$

Inductances for steel-core aluminium conductors are given in Table 1, and for copper conductors in Table 2. It will be noted that the values vary within quite narrow limits, and if conductor sizes and spacings are selected in reasonable relation to line voltage, the variation is even smaller. For example, the value of L' in Table 3 should be noted.

By making $l = 1$ in eqn. (2), we obtain the resultant line reactive powers per mile.

Example.

Assume a 132 kV 3-phase 50 c/s line of the type indicated in Table 3.

Then $L' = 0.0021$ henry and eqn. (2) becomes

$$\begin{aligned} (Q_L - Q_C)/\text{mile} &= 0.01884 \times 50 \left(0.0021 I^2 - \frac{132^2}{0.0021 \times 10^5} \right) \\ &= 0.00198 I^2 - 78.2 \text{ resultant line kVAr per mile} \end{aligned}$$

From this the resultant line reactive power per mile can easily be found for any value of I . If $I = 0$, it is -78.2 kVAr/mile, which indicates that the charging power of the line is 78.2 kVAr/mile. When the result of eqn. (2) is positive, it indicates lagging reactive

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Coventry is with British Insulated Callender's Cables Ltd.

Table 1

VALUES OF L' (IN HENRYS PER MILE) FOR STEEL-CORE ALUMINIUM CONDUCTORS

Conductor size (copper equivalent)	Equivalent spacing							
	10 ft	12 ft	14 ft	16 ft	18 ft	20 ft	22 ft	24 ft
in ²								
0.06	0.002 29	0.002 36	0.002 41					
0.1	0.002 20	0.002 26	0.002 32	0.002 35				
0.15	0.001 95	0.002 01	0.002 06	0.002 10	0.002 14			
0.2	0.001 89	0.001 96	0.002 01	0.002 05	0.002 08	0.002 12		
0.25		0.001 93	0.001 98	0.002 01	0.002 05	0.002 08	0.002 12	
0.3			0.001 95	0.001 99	0.002 02	0.002 06	0.002 09	0.002 12
0.4				0.001 94	0.001 98	0.002 02	0.002 04	0.002 07

Table 2

VALUES OF L' (IN HENRYS PER MILE) FOR COPPER CONDUCTORS

Conductors		Equivalent spacing						
Size	Construction	1 ft	2 ft	3 ft	4 ft	6 ft	8 ft	10 ft
in ²								
	No. 8 s.w.g.	0.001 69	0.001 92	0.002 05	0.002 14	0.002 28	0.002 28	0.002 44
	No. 4 s.w.g.	0.001 60	0.001 82	0.001 96	0.002 04	0.002 17	0.002 27	0.002 34
0.025	3/0.104	0.001 62	0.001 85	0.001 98	0.002 07	0.002 20	0.002 29	0.002 36
0.04	3/0.131	0.001 55	0.001 77	0.001 90	0.002 00	0.002 13	0.002 22	0.002 29
0.06	7/0.116	0.001 46	0.001 69	0.001 82	0.001 91	0.002 04	0.002 14	0.002 21
0.075	7/0.118	0.001 46	0.001 68	0.001 81	0.001 91	0.002 04	0.002 14	0.002 21
0.1	7/0.136	0.001 41	0.001 64	0.001 77	0.001 86	0.001 99	0.002 08	0.002 16
0.15	7/0.168	0.001 35	0.001 58	0.001 70	0.001 79	0.001 93	0.002 02	0.002 09
0.20	19/0.116	0.001 29	0.001 51	0.001 64	0.001 73	0.001 86	0.001 96	0.002 02
0.25	19/0.131	0.001 25	0.001 47	0.001 60	0.001 70	0.001 83	0.001 92	0.001 99

Table 3

Line voltage	S.C.A. conductor (copper equivalent)	Equivalent spacing	L'
kV	in ²	ft	henry
66	0.1	10	0.00220
88	0.15	12	0.00201
132	0.175	17	0.00209
165	0.25	19	0.00206
220	0.4	21	0.00202

power, i.e. the line is acting as a reactive load. When the result is negative, it indicates leading reactive power, i.e. the line is acting as a reactive source. When it is zero, the line will be neither generating nor consuming reactive power, and for this condition

$$L'I^2 = \frac{(kV)^2}{L' \times 10^5}$$

Hence

$$I = \frac{kV}{316L'} \quad \dots \dots \dots (3)$$

This is the current of natural or surge-impedance load, assuming this load to have unity power factor.

In the above example, I_n (current of natural load) is 199 amp.

In Fig. 1, the resultant line reactive powers per mile of all the lines indicated in Table 3 have been plotted against current. The curves bring out clearly the difference in reactive-power performance between the lower- and higher-voltage lines. The 66 kV line becomes a reactive load when transmitting all currents

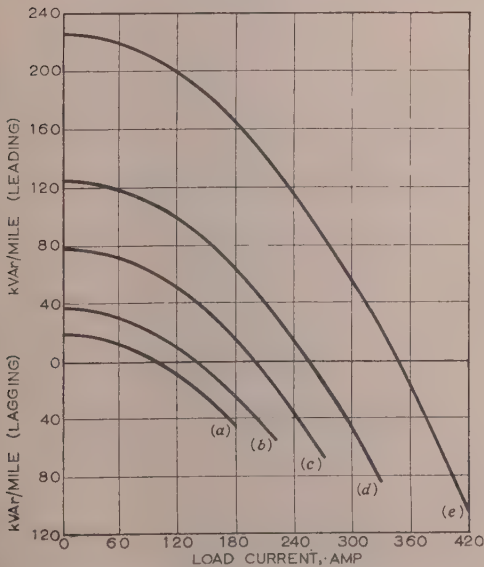


Fig. 1.—Resultant reactive-power curves of 3-phase single-circuit overhead lines.

- (a) 0.1 in² 66 kV.
 (b) 0.15 in² 88 kV.
 (c) 0.175 in² 132 kV.
 (d) 0.25 in² 165 kV.
 (e) 0.4 in² 220 kV.

The conductors are steel-core aluminium, and the sizes are copper equivalent (see Table 3).

over 100 amp; the 220 kV line does not become a reactive load until the line current is over 345 amp. The charging reactive power per mile of the former is 18.63 kVAr/mile and of the latter 226 kVAr/mile.

Owing to the small variation that occurs in L' with all practical conductor sizes and spacings, the curves in Fig. 1 can be used to find the reactive-power effect of any line operating at the voltages considered, with a fair degree of accuracy.

(3) LINES CONTAINING SERIES CAPACITORS

For lines containing series capacitors, eqn. (2) can be rewritten as follows:

$$Q_L - Q_C = 0.01884f \left[I^2 \left(L' - \frac{1}{39.6f^2 C_s} \right) - \frac{I(kV)^2}{L' \times 10^5} \right] \quad (4)$$

where C_s = Total capacitance per phase of the series capacitor, farads.

It will be noted that the effect of the series capacitor depends on the line current, and therefore will not increase the charging reactive power. It will, however, increase the line current at which the line changes from a reactive-power source to a reactive load, and when it is large enough to compensate (or more than compensate) the reactance of the line, it will ensure that the line never becomes a reactive load but always remains a reactive source.

(4) LOW-VOLTAGE LINES

In all lines of voltages of 11 kV and below, the charging reactive power is negligible and consequently the second term inside the bracket of eqn. (2) can be neglected. The formula then becomes

$$Q_L = 0.01884f/L'I^2 \quad (5)$$

and per mile at 50 c/s it becomes

$$Q_L/\text{mile} = 0.942L'I^2$$

As currents in low-voltage mains supplies can be of the same order as in high-voltage lines, and as the counterbalancing charging reactive power is practically absent, the resultant line reactive power per mile of low-voltage lines can be high and will, of course, always be a reactive load.

Example.—400-volt 3-phase 50 c/s line; conductor 0.15 in²; 1 ft spacing; current 200 amp; $L' = 0.00135$ henry (from Table 2).

Resultant line reactive power = $0.942 \times 0.00135 \times 200^2 = 50.9$ kVAr/mile.

Low-voltage lines are, of course, necessarily short, but this is usually counterbalanced by their number, and their total effect on a system may be appreciable.

(5) ACKNOWLEDGMENTS

The author wishes to express his thanks to British Insulated Callender's Cables, Ltd., for permission to publish the paper, and to his colleagues, Messrs. W. G. Hawley, R. O. M. Powell and C. W. Nixon, who have helped him by their suggestions. Mr. C. W. Nixon also gave valuable assistance by making the calculations necessary to the preparation of the Appendix.

(6) APPENDIX

(6.1) Accuracy of the Method

If we consider specific examples of overhead lines which are electrically long, and calculate their reactive-power performance for varying load currents, both by the author's method and by the

method of rigorous solution, and if we plot the results, we can see clearly the degree of any divergency that the two methods may exhibit. This has been done in Fig. 2 for three such lines—one

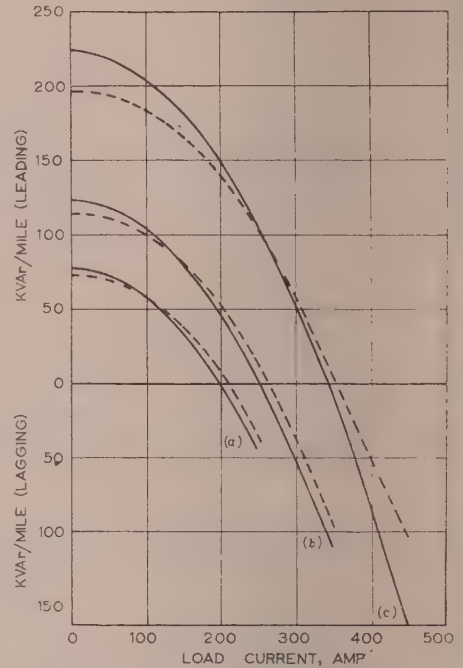


Fig. 2.—Resultant reactive-power curves of "long" overhead lines.

- (a) 0.175 in² 132 kV 130-mile single-circuit line.
- (b) 0.25 in² 165 kV 180-mile single-circuit line.
- (c) 0.4 in² 220 kV 250-mile single-circuit line.

— Author's method.

- - - Rigorous method.

The conductors are steel-core aluminium and the sizes are copper equivalent (see Table 3).

being a 220 kV 250-mile line with a 0.4 in conductor, another being a 165 kV 180-mile line with 0.25 in² conductor, and the third being a 132 kV 130-mile line with 0.175 in² conductor (see Table 3). The full-line curves represent the author's method, and the dotted-line curves represent the rigorous method.

It will be noted that in every case the full-line and dotted-line curves cross each other, and it may be assumed that this will happen with any other example of a long line. For the particular load currents at which the curves cross, the two methods give results which are in exact agreement. The further we move away in either direction from these currents, the greater the divergency exhibited by the two methods. At the extremes of the current ranges, the divergency is serious for the example considered, particularly for the 220 kV example. However these lines are electrically long, and the rigorous results are theoretical and would be modified in practice. This will be discussed later.

If we consider lines which are electrically short—say not more than 100 miles in length—we find that the divergencies, even at the extremes of the current ranges, are not important. For example, if we take the 220 kV example referred to above, but assume a length of 100 miles instead of 250 miles, we obtain the comparison shown in Table 4.

From Table 4, it should be borne in mind that 450 amp is high current for a 220 kV line. It is about 100 amp above the current of the surge-impedance load, and the divergency shown will decrease with decrease of current.

Let us return to electrically long lines, and consider the 220 kV

Table 4

RESULTANT REACTIVE POWER PER MILE OF 220 kV OVERHEAD
LINE WITH 0.4 IN² STEEL-CORE ALUMINIUM CONDUCTOR

Load current	Author's method	Rigorous method	
		100 miles long	150 miles long
amp.	kVAr	kVAr	kVAr
0	226	-220	-214.5
450	160	150	141.5

50-mile example. The rigorous solution indicates that the voltage at the sending end, with the receiving-end voltage fixed, will be 200 kV at zero load current, and 256 kV when the load current is 450 amp. (It is assumed that the power factor of the load is unity.) It is this big voltage swing, due mainly to the capacitance and inductance of the line, which is chiefly responsible for the differences between the results obtained by the author's method and those obtained rigorously. In practice, however, this swing would not be tolerated and would be reduced by use of equipment such as synchronous condensers connected to the receiving end of the line. This equipment would draw reactive power from the line during periods of light load and inject reactive power into it during periods of heavy load.

In our example, if a synchronous condenser is connected at the receiving end and it draws a current, lagging 90° (approximately), equivalent to 75 amp at 220 kV, when the load current is zero, the voltage at the sending end will be very close to 220 kV, and the current at the sending end will be 74.5 amp leading 90° (approximately). The resultant reactive power of the combination of line and synchronous condenser will be 28 400 kVAr (leading). The reactive load of the synchronous condenser is 8 570 kVAr (lagging); and therefore the reactive-power output

of the line alone will be $28\,400 + 28\,570 = 56\,970$ kVAr (leading), which, dividing by 250, gives an output per mile of 228 kVAr (leading). This result, according to the author's method, would be 226 kVAr (leading), i.e. a very close agreement.

Similarly, if we take the case when the load current is 450 amp at unity power factor, and inject by means of the synchronous condenser the equivalent of 60 amp at 220 kV, leading 90° (approximately), the voltage and current at the sending end will be 242 373 volts $\angle 31^\circ 43'$ and 455.8 amp $\angle 26^\circ 35'$ respectively, and the resultant reactive power of the combination of the line and synchronous condenser will be 16 900 kVAr (lagging). The reactive-power input of the synchronous condenser is 22 880 kVAr (leading); therefore the resultant reactive power of the line alone will be $16\,900 + 22\,880 = 39\,780$ kVAr (lagging) which, dividing by 250, gives a line load per mile of 159 kVAr (lagging). This result, according to the author's method, would be 160 kVAr (lagging), again a very close agreement.

The synchronous-condenser currents selected above are reasonable in the circumstances because they result in a voltage regulation of 10%, as compared with 25% in the unmodified line, but a load current of 450 amp is an extreme case because, among other things, it causes in a 250-mile line an angular displacement of the voltage of about 30°, which may be too high for stability. It was chosen because, if the author's method can be shown to be reasonably accurate at the extremes of the current range, it will also be accurate at all intermediate currents.

In the paper, load currents have been assumed to be in phase with the voltage. Synchronous condensers are often used to bring about this condition, and if they are used as well to inject leading reactive power into the line as indicated above, they may have to be very large. This can be avoided by using them in association with static capacitor banks, which can be arranged to take care of the lagging reactive component of the load, leaving the synchronous condenser to deal only with the line reactive power.

RELUCTANCE OF THE TEETH OF A SLOTTED ARMATURE

By S. NEVILLE, B.Sc.(Eng.), Member.

(The paper was first received 7th October, and in revised form 15th December, 1955. It was published as an INSTITUTION MONOGRAPH in April, 1956.)

SUMMARY

The paper treats of a practical method for calculating the m.m.f. required for a tapered tooth, which is correct over the whole range of flux densities and for all proportions of teeth, slots or ducts. Working curves are given for a typical core steel.

INTRODUCTION

Fifty-five years ago Hird¹ gave an accurate method of calculating the m.m.f. required for a tapered tooth, embodying correct allowance for the air space in an adjacent parallel slot; but the original treatment was inapplicable where there was also some non-magnetic space in the axial length of the core, such as ventilating ducts or the insulation between stampings. Subsequently, however, Carter² pointed out that Hird's method could be correctly applied to such cases if the tooth was treated as made of a composite material—part iron, part space—of which the magnetic properties were known and were uniform throughout. The purpose of this note is to present Carter's adaptation of Hird's method in a form familiar to designers and convenient in practical use. It covers accurately the whole range of flux densities, and all proportions of teeth, slots or ducts; it fails only for an entirely parallel tooth, the result then becoming indeterminate.

(1) THEORETICAL BASIS

With reference to Fig. 1, it is necessary to distinguish between the following estimates of the flux density at any point in the tooth:

- B' = Apparent flux density in the iron, assuming the total flux of the whole slot pitch, ABCD, to pass through the iron.
 B'' = Fictitious flux density in the iron, assuming the whole flux of the tooth region, ABXY, to pass through the iron.
 B = Real flux density in the iron, in terms of which the necessary magnetizing force H is known.

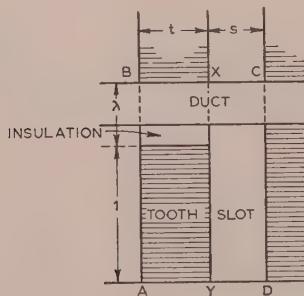


Fig. 1.—Cross-section of tooth, slot and duct.
Definition of B , B' , B'' .

These three values are in descending order of magnitude. The excess of B'' over B allows for flux in the insulation of stampings and in radial ducts, within the width of the tooth; the excess of

B' over B'' allows for flux in the slot, over the whole axial length, B' being the density ordinarily appearing on a design calculation sheet, and from which it is required to calculate the tooth m.m.f.

At any peripheral cross-section, for a real density B (gauss) in the iron, with corresponding magnetizing force H (oersteds), there will be a flux density in all adjacent spaces which is numerically equal to the magnetizing force.

If λ is the total non-magnetic axial length expressed as a fraction of the net length of iron, and if the additional flux in insulation and ducts, within the width of the tooth, is assumed to pass through the net iron, the density therein will be raised by λH . Thus the 'fictitious' density in the iron is

$$B'' = B + \lambda H \quad (1)^*$$

Similarly, if the additional flux in the slot (of width, say s), over the whole axial length, is assumed to pass through the tooth region (of width, say t), the average density therein will be raised by $(s/t)H$; and if it is further assumed to pass only through the iron, the density therein will be raised by $(s/t)(1 + \lambda)H$. Thus the 'apparent' density in the iron is

$$B' = B'' + (s/t)(1 + \lambda)H \quad (2)$$

In a straight-tapered tooth of radial length L , if t_1 is the width at the wider end, t_2 that at the narrower end, and t that at any section distant l from the narrower end, we have from eqn. (2)

$$tB' + s(1 + \lambda)H = tB'' = t_2B'_2$$

whence

$$t = t_2B'_2/B' - s(1 + \lambda)H/B'$$

or

$$dt = t_2B'_2d(1/B') - s(1 + \lambda)d(H/B')$$

Then, since $dt/dl = (t_1 - t_2)/L$, the average magnetizing force over the length of the tooth is

$$\begin{aligned} H(\text{mean}) &= \frac{1}{L} \int_0^L H dl = \frac{1}{t_1 - t_2} \int_{t_2}^{t_1} H dt \\ &= \frac{t_2}{t_1 - t_2} B'_2 \int_{B'_2}^{B'_1} H d(1/B') - \frac{s(1 + \lambda)}{t_1 - t_2} \int_{B'_2}^{B'_1} H d(H/B') \\ &= \frac{t_2}{t_1 - t_2} B'_2 (\beta_2 - \beta_1) + \frac{s}{t_1 - t_2} (\alpha_2 - \alpha_1) \quad (3) \end{aligned}$$

where β_2 , β_1 , and α_2 , α_1 , are the values at B'_2 and B'_1 , respectively, of the integrals

$$\beta = - \int H d(1/B')$$

and

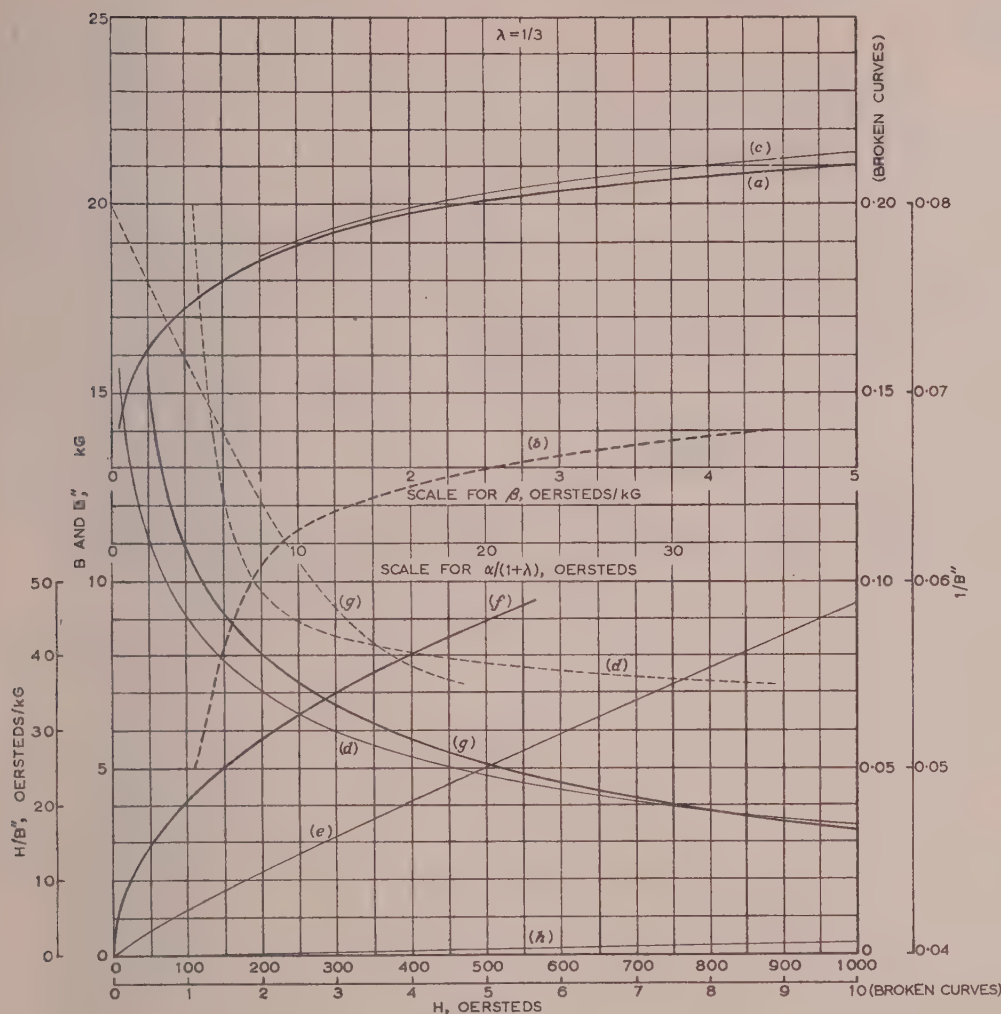
$$\alpha = (1 + \lambda) \int H d(H/B')$$

the relation of B'' to H being given by eqn. (1). The change of sign between the two terms in eqn. (3) should be noticed.

Application of the method requires the preparation of curves for β and α , derived by graphical integration from the original

* Correspondence on Monographs is invited for consideration with a view to publication.
Mr. Neville is with the Metropolitan-Vickers Electrical Co., Ltd.

* With rationalized M.K.S. units, densities being in webers per square metre, H would be replaced throughout by $4\pi/10 \times 10^{-6}$ ampere-turn per metre.

Fig. 2.—Derivation of integral curves from given B/H curve.

$$\beta = - \int H d(1/B'')$$

$$\alpha = (1 + \lambda) \int H d(H/B'')$$

- (a) Original curve of B to a base of H .
 (b) As at (a) but with enlarged scale of H .
 (c) B'' plotted to the base H .
 (d) Deduced curves of $1/B''$ to base H , using two scales of $1/B''$, shown by right-hand margin, and the two scales of H .

- (e) Deduced curve of H/B'' plotted to base H .
 (f) $\alpha/(1 + \lambda)$ plotted against H/B'' .
 (g) β plotted against the two different scales of $1/B''$, and with the scale of β enlarged tenfold for the lower part.
 (h) Flux density equal to λH , plotted to base of H .

magnetization curve of the material and a value of λ appropriate to the work in view. The lower limit of integration is arbitrary and need not be known. The curves can be extended to densities below the lower limit of integration if required; the values of β and α simply become negative.

(2) CONSTRUCTION OF WORKING CURVES

Fig. 2 is an example of the graphical work, for a variety of low-silicon core steel, λ being taken as $1/3$ —i.e. (gross core length)/(net iron length) = $4/3$. The original curve of B to a base of H is (a); the lower part, curve (b), is drawn to an enlarged scale of H . Curve (c), obtained by adding $H/3$ to the values of B , gives B'' to a base of H . The deduced curves of $1/B''$ and H/B'' plotted to the base H are shown by (d) and (e); the former has been plotted in two parts, using the two scales of $1/B''$ shown at the right-hand margin, and the two scales of H , for densities above and below 14 kG respectively. The lower

limit of integration has been taken as 5 kG; the integrations therefore commence at $1/B'' = 0.2$ and H/B'' practically zero. The H/B'' curve being nearly straight is readily integrated by ordinates; the $1/B''$ curves require a planimeter. The resulting values of $\alpha/(1 + \lambda)$ and of β [see curves (f) and (g)] are plotted against the vertical scales of H/B'' and $1/B''$ respectively, the latter in two parts, using the two different scales of $1/B''$ and also an enlarged scale of β for the lower part. For any value of β , reference to the curve of $1/B''$ gives the corresponding value of B'' , which is then re-plotted in Fig. 3, to a base of β , where it may be seen as the curve marked $s/t = 0$; its lower part is plotted to an enlarged scale of β and is shown by the broken curve. It should be noted that the β values need to be read accurately, since they are multiplied by B'' . The α values could, of course, be shown in terms of B'' by separate curves; but they are more conveniently shown against the corresponding values of β , by the separate non-uniform scale at the top.

It is unnecessary to extend the graphical integrations beyond

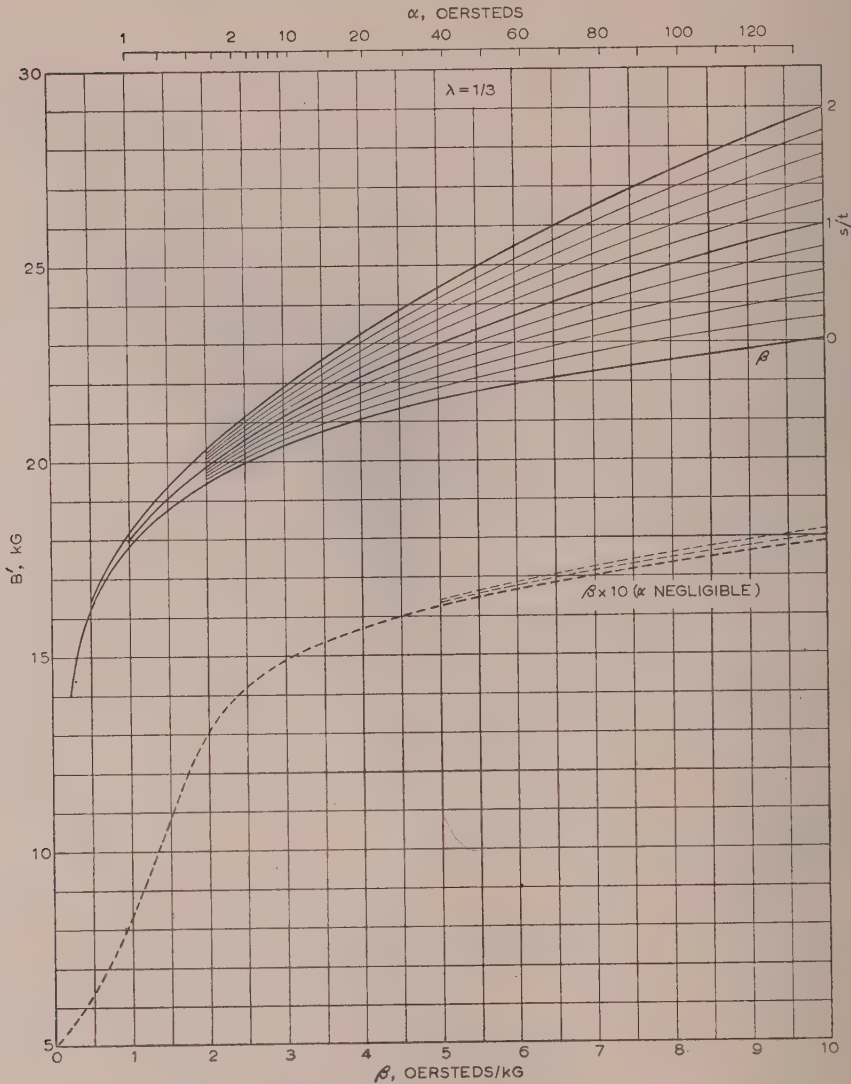


Fig. 3.—Working curves for material of Fig. 2.

$$H_{mean} = \frac{t_2}{t_1 - t_2} B'_2 (\beta_2 - \beta_1) + \frac{s}{t_1 - t_2} (\alpha_2 - \alpha_1)$$

the saturation point, which in this material occurs at $B = 21.1$ kG, $H = 1000$ oersteds, and $B_0 = 20.1$ kG. At higher densities $B = B_0 + H$; and therefore $B'' = B_0 + (1 + \lambda)H$, or $H = (B'' - B_0)/(1 + \lambda)$. Thus

$$\begin{aligned} \int Hd(1/B') &= \frac{1}{1 + \lambda} \int (B'' - B_0)d(1/B'') \\ &= -\frac{1}{1 + \lambda} \left(\log_e B'' + \frac{B_0}{B''} \right); \end{aligned}$$

$$\text{and } \beta_2 - \beta_1 = - \int_{B'_1}^{B'_2} Hd(1/B'') = \frac{1}{1 + \lambda} \left(\log_e \frac{B'_2}{B'_1} + \frac{B_0}{B'_1} - \frac{B_0}{B'_2} \right) \quad (4a)^*$$

* With rationalized M.K.S. units this expression would be multiplied by $10^7/4\pi$; eqn. (4b) would not be affected.

$$\text{Similarly, } \int Hd(H/B'') = \left(\frac{1}{1 + \lambda} \right)^2 B_0 \left(\log_e B'' + \frac{B_0}{B''} \right)$$

$$\text{and } \alpha_2 - \alpha_1 = (1 + \lambda) \int_{B'_1}^{B'_2} Hd(H/B'') = (\beta_2 - \beta_1) B_0 \quad (4b)$$

In Fig. 3 the curves have been extended beyond $\beta = 4.6$ and $\alpha = 30$ in this way. For practical convenience a set of curves giving apparent density B' for a series of values of s/t up to 2.0 is constructed by adding the values of $(s/t)H$ to the bottom curve; the appropriate values of H are obtained by reference back to the curve of B'' in Fig. 2.

Although a particular value of λ has been assumed, there is no perceptible error in applying the resulting curves to a class of work in which λ' actually has a somewhat different value, provided that an artificial adjustment is made in the slot width, so

that $(1 + s'/t)(1 + \lambda')$ remains correct (at the narrow end of the tooth); i.e. if $(s' + t_2)$ is modified inversely as $(1 + \lambda')$.

For a solid core, such as a turbo-generator rotor, λ provides only for radial ducts; where there are no ducts $\lambda = 0$, and Hird's original method is a complete treatment.

(3) NUMERICAL EXAMPLE

A tooth 7.65 in long tapers in width from 1.84 in to 1.02 in, the slot being 0.89 in wide; the gross core length is 95 in, and the net iron length is 70 in. Find the total m.m.f. required for an apparent flux density of 24 kG at the narrow end of the tooth:

Since the actual value of λ' is given by

$$\lambda' = 95/70 - 1 = 0.357$$

$t_2 + s = 1.91$ in must be increased to

$$t_2 + s' = 1.91 \times 1.357/1.333 = 1.945 \text{ in}$$

making $s' = 0.925$.

Then we have, referring to Fig. 3,

$t_2 = 1.02$	$s'/t_2 = 0.905$	$B'_2 = 24.0$	$\beta_2 = 6.70$	$\alpha_2 = 72$
$t_1 = 1.84$	s'/t_1 not used.	$B'_1 = 13.3$	$\beta_1 = 0.21$	$\alpha_1 = 0$
$t_1 - t_2 = 0.82$		$\beta_2 - \beta_1 = 6.49$	$\alpha_2 - \alpha_1 = 72$	

$$H(\text{mean}) = \left(\frac{1.02}{0.82} \times 24.0 \times 6.49 \right) + \left(\frac{0.925}{0.82} \times 72 \right)$$

$$= 194 + 81 = 275 \text{ oersteds.}$$

A carefully plotted curve of the values of H calculated at eight sections of the tooth confirms this value.

$$\text{Thus total ampere-turns} = 275 \times 7.65 \times 2.54 \times 10/4\pi = 4250.$$

(4) ACKNOWLEDGMENT

The paper was prepared in the Plant Department of the Metropolitan-Vickers Electrical Co., Ltd., and the author wishes to thank Mr. H. West, Director and Chief Electrical Engineer, for permission to publish it.

(5) REFERENCES

- (1) HIRD, W. B.: 'Reluctance of the Teeth of a Slotted Armature', *Journal I.E.E.*, 1900, **29**, p. 936.
- (2) CARTER, F. W.: 'The Magnetic Field of the Dynamo-Electric Machine', *ibid.*, 1926, **64**, p. 1120.

THE USE OF THE BLACKBURN A.C. NETWORK ANALYSER IN THE ANALYSIS OF POWER SYSTEM FAULTS

By J. H. BANKS, M.Sc., and K. C. PARTON, B.Sc., Graduates.

(The paper was first received 23rd September, 1955, and in revised form 28th December, 1955. It was published as an INSTITUTION MONOGRAPH in April, 1956.)

SUMMARY

The paper describes the use of the Blackburn a.c. network analyser in analytical and analogue methods of investigating faults on 3-phase power systems. In the analytical approach, after defining the faults in matrix form by symmetrical components, the ensuing equations are solved using the Blackburn analyser without the need for further matrix manipulation.

In the analogue method, it is shown how the constraints imposed upon the phase voltages and currents can be applied directly to the sequence networks by the connection of symmetrical-component operational matrices at the fault positions. No knowledge of the relationships between the symmetrical-component currents is required, but it is necessary to know the relationships existing between the phase voltages and currents by virtue of the applied faults.

The 'equivalent series' connections of the sequence networks, derived recently by Kirschbaum in America for application to the conventional-type network analyser, are shown to be unnecessary when using the Blackburn machine.

Finally, systems containing unsymmetrical simultaneous faults are investigated using both methods.

LIST OF SYMBOLS

- V, I, Z = Voltage, current and impedance matrices, respectively, of the system in terms of phase quantities.
 V', I', Z' = Voltage, current and impedance matrices, respectively, in terms of symmetrical-component parameters of the system with balanced faults.
 V'', I'', Z'' = Voltage, current and impedance matrices, respectively, in terms of symmetrical-component parameters of the system with unbalanced faults.
 C = Fault-constraint connection matrix.
 A = Symmetrical-component connection matrix.
 $0, 1, 2$ = Subscripts denoting zero, positive and negative phase-sequence components respectively.
 $a = -0.5 + j0.866$.

(1) INTRODUCTION

Whilst the prime functions of any analogue machine are to simulate electrical and physical problems in model form, its value is greatly enhanced if it may also be used as a mathematical tool in analytic investigations. The Blackburn type of a.c. network analyser possesses this facility and for many problems, therefore, has distinct advantages over the conventional form of analyser which employs variable-impedance elements. When applying the Blackburn analyser¹ to interconnected-network problems, 'ideal' voltage transformers are used to represent the various resistive, inductive and capacitive components of the network. Negative resistors, resistance-free inductors and current sources can be represented without difficulty by single universal units of the analyser, whilst two units coupled together

serve to represent mutual inductors, network transformers and other coupling effects. When employing the machine as a mathematical device for the solution of simultaneous equations, for instance, the voltage transformers are interconnected to form the various coefficients of the unknown variables.

The application of the Blackburn analyser to the solution of fault investigations in 3-phase power systems may entail the use of both its analogue and mathematical facilities. For instance the most general type of series fault^{2,6} which can occur in a 3-phase transmission system is one in which three unequal impedances are introduced in the three line conductors between two points in the system. There is no simple way of interconnecting the sequence networks directly to meet the constraints imposed by such a condition, since the mutual-impedance elements in the network equations of performance are non-bilateral.

It is possible to overcome this complication in two ways. The first is to define the equations of the faulted network completely, and to solve these either on the Blackburn analyser or by an automatic digital computer. The second is to interconnect the sequence networks by means of symmetrical-component operational matrices to form the phase values of voltage and current and then to constrain these phase quantities by further interconnection to fulfil the fault condition. Both methods are described in Section 2, and their application is considered in Section 3.

If the series fault is symmetrical with respect to the reference phase, the mutual-impedance elements are bilateral, and solution is possible by classical analogue and analytic methods. For those cases of fault in which the real parts of the impedances of two of the line conductors exceed the real part of the third line conductor impedance, negative-resistance elements appear in the interconnected sequence network and representation of these cases is not easily possible by a conventional a.c. network analyser; no difficulty is experienced when employing the transformer analogue machine, however.

The methods described herein may be applied to all types of fault, including simultaneous dissymmetries. Since they are concise in principle they may be set up with ease when using a large Blackburn analyser. For smaller analysers, the statement of the problem in matrix terms may first be condensed by the evaluation of matrix products.

Although the paper is a statement of principle, experimental experience has indicated that errors incurred when employing the analogue form of solution on the Blackburn machine will be within $\pm 2\%$ of base quantities, for all but the largest networks.

Larger errors may be met when evaluating the matrix equations if the component equations are ill-conditioned.

(2) ANALYSER AND MATRIX TECHNIQUES

(2.1) Transformer Analogue Representation

A full description of the principle and working of the Blackburn analogue machine can be found elsewhere,¹ and it will suffice here to mention the basic principle.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Banks is with the Central Electricity Authority, and was formerly with The General Electric Co., Ltd.

Mr. Parton is with The General Electric Co., Ltd.

If a voltage x is applied to an ideal transformer tapped in the ratio of $1/y$ the output voltage obtained will be of magnitude yx . This was developed by Blackburn, using two transformers and two voltages to represent complex quantities, in order to multiply the voltage $x + jx'$ by a constant $y + jy'$ to give the output voltage $(yx - y'x') + j(yx' + y'x)$. By reversing the polarity of the tapped windings, it is possible to make either y or y' negative at will, permitting the facile representation of any impedance $\pm A \pm jB$.

Where analyser transformer circuits are shown in this paper they will in general be drawn for non-complex quantities in order to keep the principle of the interconnections as clear as possible. Wherever non-complex multiplications are shown, however, complex multiplications may always be performed with standard Blackburn analyser units.¹

(2.1.1) Matrix Representation.

A group of complex multiplier and mutual units may be used to obtain the solution of a set of simultaneous equations with complex coefficients.

Consider the simultaneous equations

$$\begin{cases} V_1 = Z_{11} \cdot I_1 + Z_{12} \cdot I_2 \\ V_2 = Z_{21} \cdot I_1 + Z_{22} \cdot I_2 \end{cases} \quad (1)$$

which may be written in matrix form as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad (2)$$

On the Blackburn analyser, the equations would be represented by the circuit of Fig. 1, where the lower windings of each transformer have unit numbers of turns and the upper windings are tapped to the values Z_{11} , Z_{12} , etc. The settings of the mutual impedance in each row of the analyser are such as to inject appropriate voltages into the other rows. The physical arrangement of the analyser units shown in Fig. 1 thus corresponds to the transposition of the impedance matrix.

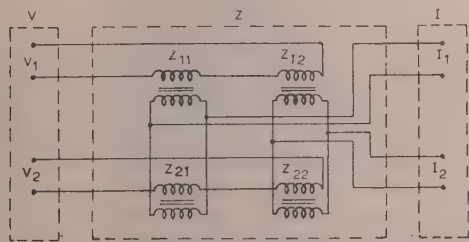


Fig. 1.—Blackburn analyser connections for solution of matrix equation of order 2.

By applying the voltages V_1 and V_2 at the left-hand terminals as shown, the voltages appearing at the right-hand terminals will be the solution for I_1 and I_2 respectively.

Generally, by applying voltages x to the left-hand column of a matrix equation circuit similar to that of Fig. 1, the resultant output voltages y on the right-hand column represent the solution of the matrix equation $x = zy$, where z may be any matrix.

If, however, the identical matrix z is used, but the voltage x is now applied to the right-hand column, the resultant voltages y' appearing on the left-hand column will represent the solution to the equation

$$y' = zx \quad \text{or} \quad x = z^{-1}y'$$

Hence, in general, any impedance or connection matrix can be used directly to operate as its matrix inverse by simply interchanging the input and output connections.

(2.1.2) Matrix Multiplication.

Matrix multiplication may be performed on the Blackburn analyser machine by direct interconnection. For example, if

$$Z = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} e & f \\ g & h \end{bmatrix} \quad (3)$$

the Blackburn circuit for the evaluation of Z will be shown as in Fig. 2. If, however,

$$Z = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} e & f \\ g & h \end{bmatrix}^{-1} \quad (4)$$

then Z will be evaluated by the interconnections of Fig. 3.

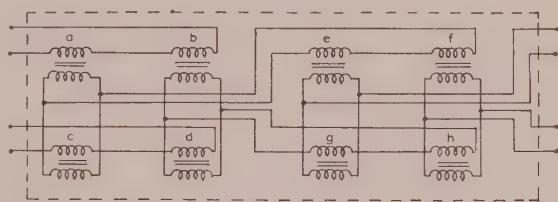


Fig. 2.—Blackburn analyser connections for matrix multiplication.

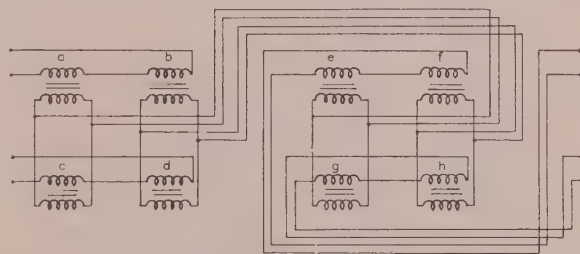


Fig. 3.—Blackburn analyser connections for multiplication of a matrix inverse by another matrix.

It is possible to simplify matrix products by combining any number of connected matrices into an equivalent matrix, with the values of the elements rapidly determined by the following procedure.

(a) After connecting together the block of matrix products, unit voltage is applied to the first element of the output column vector and zero voltages to the other elements of this vector.

(b) The voltages in all elements of the input column vector are now measured and their magnitudes are those of the elements in the first column of the equivalent single Z matrix, as shown in eqns. (5) and (6).

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{21} & Z_{22} & Z_{23} \\ Z_{31} & Z_{32} & Z_{33} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \quad (5)$$

Eqns. (14) and (15) enable the voltage and impedance matrices of the symmetrical-component representation of the unbalanced system to be calculated. The required currents may now be evaluated, either by hand, using a desk calculating machine for small problems, or by an automatic digital computer for larger problems. Alternatively, the problem can be set up and solved directly by connecting together the original matrices on the Blackburn analyser.

By considering the application of an earth fault at the terminals of a 3-phase generator, the *balanced* fault condition will be defined by the impedance matrix

$$Z' = \begin{bmatrix} Z_0 & 0 & 0 \\ 0 & Z_1 & 0 \\ 0 & 0 & Z_2 \end{bmatrix} \quad (16)$$

since the sequence networks are isolated.

The constraints imposed by the earth fault are such that

$$\begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot I_1 \quad (17)$$

so that the connection matrix is given by

$$C = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (18)$$

Then the faulted impedance and voltage matrices are given by

$$Z'' = \bar{C}_t Z' C = Z_0 + Z_1 + Z_2 \quad (19)$$

$$V'' = \bar{C}_t V' = V_1 \quad (20)$$

(2.4) Blackburn Treatment of Kron Constraint Approach

By rewriting $V'' = \bar{C}_t V'$ in the form $V' = (\bar{C}_t)^{-1} V''$, the block diagram of the connections needed for solution of Section 2.3 on the Blackburn analyser is shown in Fig. 6. If the symmetrical-

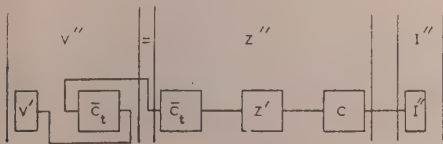


Fig. 6.—Block diagram of analyser interconnections for the solution of any fault expressed in terms of the symmetrical component matrix Z' .

component voltage matrix V' is now applied, the resultant symmetrical-component fault currents I'' are given directly. This block diagram is general for all types of fault.

For the above example of a single-line-to-neutral fault, the complete Blackburn interconnections are given in Fig. 7. The voltage output V'' of the first \bar{C}_t matrix is seen to equal V_1 , and

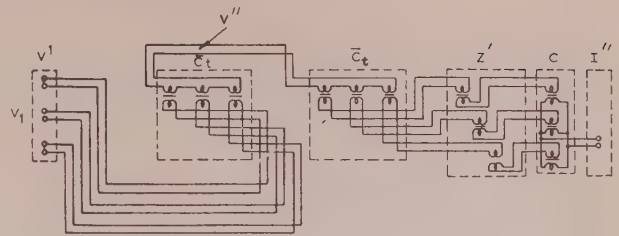


Fig. 7.—Blackburn analyser connections for matrix solution of line-to-earth fault.

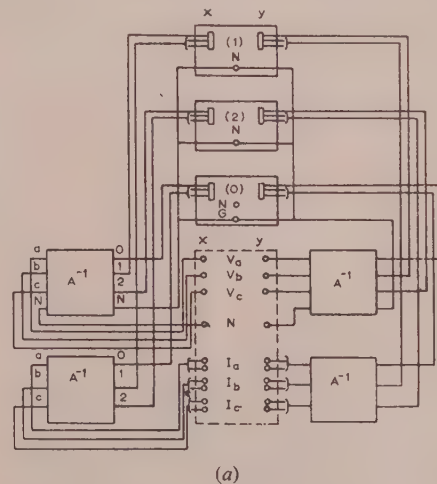
for voltage balance in the second \bar{C}_t matrix, V'' must be equal to $I''Z_0 + I''Z_1 + I''Z_2$;

$$i.e. \quad I'' = \frac{V_1}{Z_0 + Z_1 + Z_2} \quad (21)$$

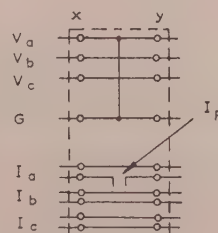
giving the requisite constraint equation for the earth fault condition.

(2.5) Blackburn Analyser Constraint Method

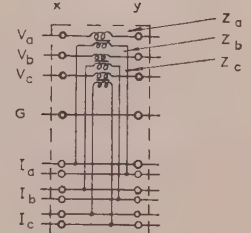
Since the Blackburn analyser can readily transform symmetrical-component quantities into phase quantities and vice versa, it is possible to analyse any type of fault dissymmetry directly without extensive knowledge of symmetrical-component manipulation or matrix techniques. The procedure may be stated as follows.



(a)



(b)



(c)

Fig. 8.—Analyser phase-constraint interconnections.

- (a) Blackburn analyser interconnections of sequence networks for any fault.
- (b) Phase constraints for application of single-line-to-earth fault.
- (c) Phase constraints for unbalanced 3-phase load.

- (a) The three sequence networks are constructed from system data and broken at the point or points of fault.
- (b) These terminations are then connected through symmetrical-component operational matrices, and the three output quantities obtained on the phase side of the symmetrical-component operators. The system of interconnection is shown in Fig. 8(a).
- (c) The constraints required on the phase quantities at the point of fault are now applied directly.
- (d) The analyser is energized and the sequence component and phase quantities are measured.

Considering the case of a single-line-to-earth fault as in Section 2.3, the phase connections required will be as shown in Fig. 8(b). As in normal Blackburn analyser procedure, the voltage leads are interconnected in similar fashion to the configuration of the actual system conductors.

Also, since the currents I_{bx} , I_{by} and currents I_{cx} , I_{cy} must be identical, they are connected together, so that the analyser voltage representing I_{bx} is the same as that for I_{by} ; similarly the voltage representing the current I_{cx} is equal to that representing I_{cy} .

The unknown phase-fault current I_{af} is given by the difference between the voltages representing the currents I_{ax} and I_{ay} .

Further analysis of the method is given in the Appendix.

(3) FAULT ANALYSIS

(3.1) Open Conductor on Line *a*

(3.1.1) Matrix Treatment.

Consider the 3-phase system of Fig. 9 shown operating normally. The equations of the network may be written down most directly using Stigant's rule⁵, which is defined briefly below.

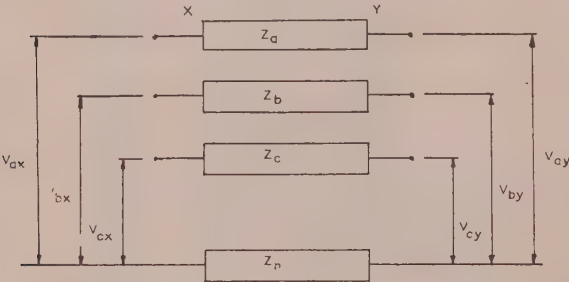


Fig. 9.—Three-phase system, operating normally.

- The paths of the independent mesh currents are chosen to suit the problem under consideration, and the elements of the impedance matrix are found as follows:
- (a) A diagonal element such as Z_{aa} is the total impedance in the path of mesh current *a*.
- (b) A non-diagonal element such as Z_{ab} is the impedance common to the mesh currents *a* and *b*. It is positive if I_a and I_b flow through the impedance in the same direction, and negative if they flow in opposite directions.
- For the system portrayed in Fig. 9, therefore, the matrix equation of performance is given by

$V_{ax} - V_{ay}$	$Z_a + Z_n$	Z_n	Z_n	I_a
$V_{bx} - V_{by}$	Z_n	$Z_b + Z_n$	Z_n	I_b
$V_{cx} - V_{cy}$	Z_n	Z_n	$Z_c + Z_n$	I_c

$$= \begin{bmatrix} Z_a + Z_n & Z_n & Z_n \\ Z_n & Z_b + Z_n & Z_n \\ Z_n & Z_n & Z_c + Z_n \end{bmatrix} \cdot \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (22)$$

Now

I_a	1	1	1	I_{a0}
I_b	1	a^2	a	I_{a1}
I_c	1	a	a^2	I_{a2}

$$= \begin{bmatrix} 1 & 1 & 1 \\ 1 & a^2 & a \\ 1 & a & a^2 \end{bmatrix} \cdot \begin{bmatrix} I_{a0} \\ I_{a1} \\ I_{a2} \end{bmatrix} \quad (23)$$

Substituting for (I_a , I_b , I_c) in eqn. (22), and multiplying through out by A^{-1} , the equivalent symmetrical-component equation is

$V_{ax0} - V_{ay0}$	$Z_a + Z_b + Z_c + 9Z_g$	$Z_a + a^2Z_b + aZ_c$	$Z_a + aZ_b + a^2Z_c$	I_{a0}
$V_{ax1} - V_{ay1}$	$Z_a + aZ_b + a^2Z_c$	$Z_a + Z_b + Z_c$	$Z_a + a^2Z_b + aZ_c$	I_{a1}
$V_{ax2} - V_{ay2}$	$Z_a + a^2Z_b + aZ_c$	$Z_a + aZ_b + a^2Z_c$	$Z_a + Z_b + Z_c$	I_{a2}

$$= \frac{1}{3} \begin{bmatrix} Z_a + Z_b + Z_c + 9Z_g & Z_a + a^2Z_b + aZ_c & Z_a + aZ_b + a^2Z_c \\ Z_a + aZ_b + a^2Z_c & Z_a + Z_b + Z_c & Z_a + a^2Z_b + aZ_c \\ Z_a + a^2Z_b + aZ_c & Z_a + aZ_b + a^2Z_c & Z_a + Z_b + Z_c \end{bmatrix} \cdot \begin{bmatrix} I_{a0} \\ I_{a1} \\ I_{a2} \end{bmatrix} \quad (24)$$

This is the symmetrical-component equation describing the normal operating condition of the system of Fig. 9, and may be more concisely expressed as

$$V' = Z'I' \quad (25)$$

The constraint imposed by the opening of line conductor *a* is that given by the symmetrical-component equation

$$I_1 + I_2 + I_0 = 0 \quad (26)$$

The unsymmetrically faulted condition may thus be expressed in terms of two sequence components, and the relationship between the unfaulted sequence components and those of the faulted case becomes

I_0	1	0	I_0
I_1	0	1	I_1
I_2	-1	-1	

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} I_0 \\ I_1 \end{bmatrix} \quad (27)$$

$$\text{i.e.} \quad I' = CI'' \quad (28)$$

C is therefore defined for this case.

$$\text{Hence} \quad \bar{C}_t = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad (29)$$

and

$$Z'' = \bar{C}_t Z' C$$

$$= \frac{1}{3} \begin{bmatrix} 3Z_b + 3Z_c + 9Z_g & Z_b(1 - 2a + a^2) + Z_c(1 + a - 2a^2) \\ Z_b(1 - 2a^2 + a) + Z_c(1 + a^2 - 2a) & 3Z_b + 3Z_g \end{bmatrix} \quad (30)$$

The fault voltage matrix V'' , equal to $\bar{C}_t V'$, becomes

$$V'' = \begin{bmatrix} (V_{ax0} - V_{ay0}) - (V_{ax2} - V_{ay2}) \\ (V_{ax1} - V_{ay1}) - (V_{ax2} - V_{ay2}) \end{bmatrix} \quad (31)$$

$$\text{and} \quad V'' = Z''I'' \quad (32)$$

3.1.2) Blackburn Analyser Solution in Matrix Terms.

By connecting up matrix blocks as explained in Section 2.2, it is possible to solve the problem directly from a knowledge of the phase voltages, the phase impedance matrix and the constraint matrix C .

The resultant connection of matrices required in order to carry out the sequence of operations explained in Section 3.1.1 is shown in Fig. 10(a). The final A^{-1} matrix has been inserted at the

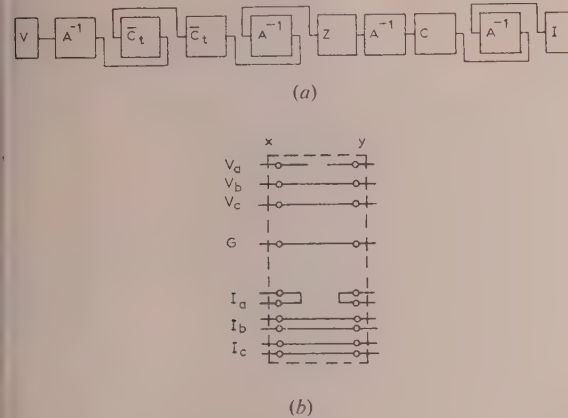


Fig. 10.—Analyser interconnections for open-conductor fault.

(a) Block diagram of matrix interconnections on Blackburn analyser for any fault expressed in terms of the phase-impedance matrix Z .

(b) Phase constraints for open-conductor fault on phase a .

and, in order that the final voltages obtained may represent directly the phase values of the fault currents on the system.

3.1.3) Direct Solution by Blackburn Analogue.

For an analogue solution, the method explained in Section 2.4 could be directly applied, the actual connections of the phase-current quantities being given in Fig. 10(b).

(3.2) Unbalanced Loading of 3-Phase System

3.2.1) Matrix Approach.

When considering the general case of the 3-phase system in Section 3.1, the non-diagonal elements of the equivalent symmetrical-component matrix equation were seen to be non-lateral. When $Z_b = Z_c$, the impedance matrix Z' becomes

$$Z' = \frac{1}{3} \begin{bmatrix} Z_a + 2Z_b + 9Z_g & Z_a - Z_b & Z_a - Z_b \\ Z_a - Z_b & Z_a + 2Z_b & Z_a - Z_b \\ Z_a - Z_b & Z_a - Z_b & Z_a + 2Z_b \end{bmatrix} \quad (33)$$

such a matrix is capable of equivalent-circuit representation, as shown in Fig. 11.

When the real part of Z_b exceeds the real part of Z_a , representation of the resulting negative resistance is not easily possible by conventional analysers, and an alternative form of equivalent circuit is necessary. Kirschbaum in America has recently advocated an equivalent series circuit to overcome this limitation, which is shown in Fig. 12, but it will be seen that considerable calculation of admittances is necessary.

When using a Blackburn analyser, however, no such transformation to the dual is necessary, since negative-resistance representation presents no difficulty, as described in Section 2,

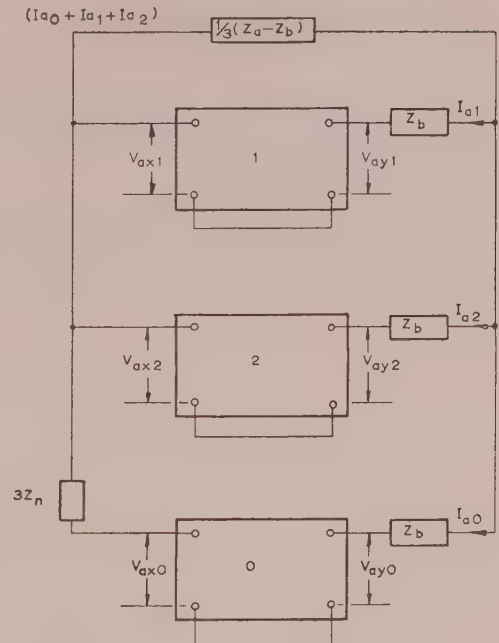


Fig. 11.—Equivalent shunt interconnections for series unbalance, for the 3-phase system of Fig. 9.

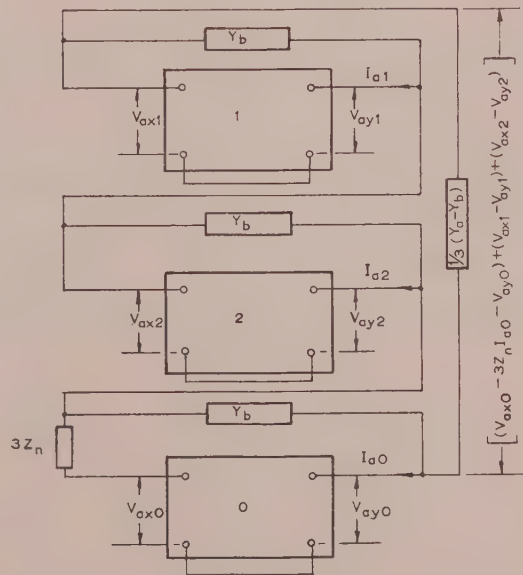


Fig. 12.—Equivalent series interconnection for series unbalance of 3-phase system.

and consequently the equivalent circuit of Fig. 11 may be used directly.

(3.2.2) Blackburn Constraint Technique.

A 3-phase unbalanced load may be directly coupled into the system in accordance with Section 2.5, and no restriction is necessary on the relative values of the 3-phase impedances, which may all be different. The scheme of the phase interconnections is shown in Fig. 8(c).

(3.3) Simultaneous Fault Condition of Open Conductor on Phase *a* and Single-Line-to-Earth Fault on Phase *b*

(3.3.1) Matrix Approach.

Consider the two-machine system shown in Fig. 13 with an open conductor on phase *a* and an earth fault on phase *b*; the impedance of the fault path is composed of the arc impedance Z_f and an earth impedance Z_g .

The balanced condition chosen is that of both machines connected with a balanced star load of phase impedance Z_f and neutral impedance Z_g at the fault point *M*. This condition is represented by Fig. 14, and using Stigant's rule, the matrix equation of the network in terms of the mesh currents shown is

$$V' = Z'I'$$

OR

0	=	$Z_{s0} + Z_f + 3Z_g$	Z_{s0}	0	0	0	0	I_{f0}
0		Z_{s0}	$Z_{s0} + Z_{t0}$	0	0	0	0	I_{s0}
V_{s1}		0	0	$Z_{s1} + Z_f$	Z_{s1}	0	0	I_{f1}
$V_{s1} - V_{t1}$		0	0	Z_{s1}	$Z_{s1} + Z_{t1}$	0	0	I_{s1}
0		0	0	0	0	$Z_{s2} + Z_f$	Z_{s2}	I_{f2}
0		0	0	0	0	Z_{s2}	$Z_{s2} + Z_{t2}$	I_{s2}

(34)

The constraint matrix is built up from the following considerations. First, the currents in the *f* mesh are subject to the constraint of a line-to-ground fault on line *b* so that

$$I_{f0} = a^2 I_{f1} = a I_2$$

Secondly, the currents in the *s* mesh are subject to the constraint of an open conductor on line *a* so that

$$I_{s0} + I_{s1} + I_{s2} = 0$$

Analysing in terms of I_{f0} , I_{s0} , and I_{s1} , the connection matrix becomes

$$C = \begin{matrix} & \begin{matrix} f_0 & s_0 & s_1 \end{matrix} \\ \begin{matrix} f_0 \\ s_0 \\ f_1 \\ s_1 \\ f_2 \\ s_2 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a & 0 & 0 \\ 0 & 0 & 1 \\ a^2 & 0 & 0 \\ 0 & -1 & -1 \end{bmatrix} \end{matrix}$$

The final impedance matrix is derived as before from the expression $Z'' = \bar{C}Z'C$. Equivalent-circuit representation is not possible as the mutual-impedance terms are non-bilateral, but the analytic solution may be obtained directly on the Blackburn analyser as indicated in Fig. 10(a). The unsymmetrically faulted

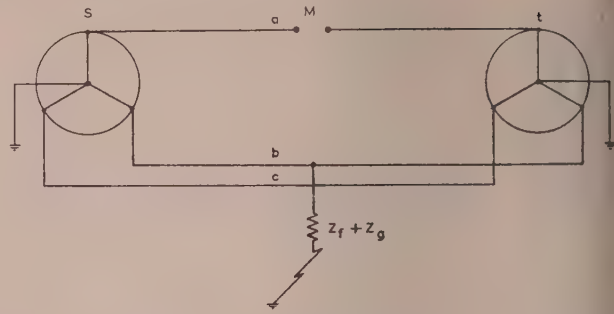


Fig. 13.—Three-phase system having an open conductor on phase *a* and a line-to-earth fault on phase *b*.

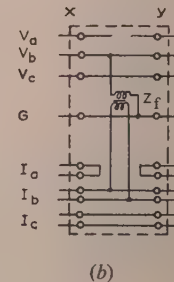
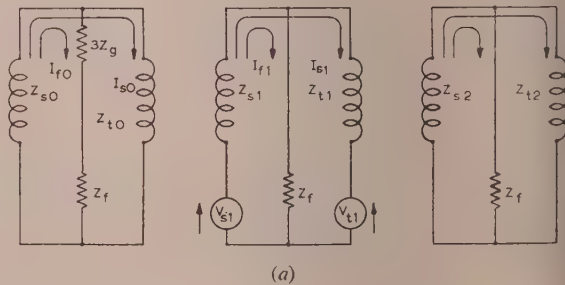


Fig. 14.—Analyser interconnections for simultaneous fault.

(a) Symmetrical-component equivalent circuit of the system of Fig. 13 with the unsymmetrical faults replaced by a balanced fault condition.
(b) Phase constraints for open conductor on phase *a* and single-line-to-earth fault on phase *b*.

impedance and voltage matrices for this condition are given below.

$$V'' = \begin{bmatrix} Z_{s0} + Z_{s1} + Z_{s2} & Z_{s0} - aZ_{s2} & a^2Z_{s1} - aZ_{s2} \\ Z_{s0} - a^2Z_{s2} & Z_{s0} + Z_{r0} + Z_{s2} + Z_{r2} & Z_{s2} + Z_{r2} \\ aZ_{s1} - a^2Z_{s2} & Z_{s2} + Z_{r2} & Z_{s1} + Z_{r1} + Z_{s1} + Z_{r2} \end{bmatrix} \quad (36)$$

$$V'' = \begin{bmatrix} a^2V_{s1} \\ 0 \\ V_{s1} - V_{r1} \end{bmatrix} \quad (37)$$

3.2) Blackburn Analogue Solution.

Following the procedure already established, the resultant interconnections required at the point of fault are as shown in fig. 14(b).

(3.4) Use of Clarke α , β , 0 Components

Whilst the analysis of unbalanced fault conditions has been performed throughout this paper using the symmetrical components of Fortescue, the α , β , 0 components of Clarke may often be applied with profit to effect a simplification in the interconnections between the component networks, to reduce the amount of phase-shift incurred, and to avoid the representation of negative resistances.

The choice of components is, however, immaterial when the analysis is performed using matrices, since no complication is created by the presence of non-bilateral mutual impedances.

(4) ACKNOWLEDGMENTS

The authors wish to express their appreciation of the encouragement they have received from Dr. E. Friedlander and Mr. J. A. Fitzpatrick of the General Electric Co., Ltd.

Acknowledgment is also made to the Directors of the Company for permission to publish the paper.

(6) REFERENCES

- (1) DAVIES, M. W. H., and SLEMON, G. R.: 'Transformer—Analogue Network Analysers', *Proceedings I.E.E.*, Paper No. 1479 S, March, 1953 (100, Part II, p. 469).
- (2) 'Westinghouse Electrical Transmission and Distribution Reference Book', p. 24.
- (3) KRON, G.: 'Tensor Analysis of Electrical Networks' (Wiley, 1939).
- (4) LEWIS, W. E., and BANKS, J. H.: 'Matrix Methods for the Evaluation of Simultaneous Faults in 3-Phase Systems', *Proceedings I.E.E.*, Monograph No. 125 S, April, 1955 (102 C, p. 231).

- (5) STIGANT, S. A.: 'The Direct Setting up of $Z_{\alpha\beta}$ for Closed Mesh Networks from the Network Diagram', *Beama Journal*, 1947, 54, pp. 28 and 65.
- (6) KIRSCHBAUM, H. S.: 'Basic Sequence Networks', *Transactions of the American I.E.E.*, 1955, Paper No. 55-278.

(7) APPENDIX

With reference to Figs. 8(a) and 8(b), the following constraints apply.

(a) Voltage constraints:

$$V_{ax} = V_{ay} = 0$$

$$V_{bx} = V_{by}$$

$$V_{cx} = V_{cy}$$

and

from which

$$V_{0x} + V_{1x} + V_{2x} = V_{0y} + V_{1y} + V_{2y} = 0$$

$$V_{0x} + a^2V_{1x} + aV_{2x} = V_{0y} + a^2V_{1y} + aV_{2y}$$

and

$$V_{0x} + aV_{1x} + a^2V_{2x} = V_{0y} + aV_{1y} + a^2V_{2y}$$

Simplifying,

$$V_{0x} = V_{0y} = V_0$$

$$V_{1x} = V_{1y} = V_1$$

$$V_{2x} = V_{2y} = V_2$$

and

$$V_0 + V_1 + V_2 = 0$$

(b) Current constraints:

$$I_{bx} + I_{by} = 0$$

and

$$I_{cx} + I_{cy} = 0$$

from which

$$I_{0x} + a^2I_{1x} + aI_{2x} + I_{0y} + a^2I_{1y} + aI_{2y} = 0$$

and

$$I_{0x} + aI_{1x} + a^2I_{2x} + I_{0y} + aI_{1y} + a^2I_{2y} = 0$$

Simplifying,

$$I_{0x} + I_{0y} = I_{1x} + I_{1y} = I_{2x} + I_{2y}$$

i.e.

$$I_0 = I_1 = I_2$$

But from the normal phase-sequence circuits used on the analyser, it follows that

$$V_0 = I_0Z_0$$

$$V_1 = I_1Z_1 - E_1$$

$$V_2 = I_2Z_2$$

But since

$$V_0 + V_1 + V_2 = 0$$

and

$$I_0 = I_1 = I_2 = I$$

therefore

$$IZ_0 + IZ_1 - E_1 + IZ_2 = 0$$

i.e.

$$I = \frac{E_1}{Z_0 + Z_1 + Z_2}$$

in accordance with known theory.

H.F. BEARING VARIATIONS ON AN ADCOCK DIRECTION-FINDER

By E. N. BRAMLEY, M.Sc., Ph.D.

(The paper was received 7th December, 1955. It was published as an INSTITUTION MONOGRAPH in April, 1956.)

SUMMARY

Measurements on a number of transmitters at ranges from 1000 to 5000 km show that the variance of snap bearings within a period of half an hour is of the order of 10 deg^2 . To reduce this variance by a factor of ten, time averaging over about 5 min is required when the bearings are taken at 10 sec intervals. Bearing changes from hour to hour and from day to day have a variance of about 1 deg^2 and thus set a limit to the error reduction obtainable by averaging over a half-hour period. The results are consistent with previously published estimates of bearing variances.

(1) INTRODUCTION

When directional observations are made on high-frequency signals from a distant transmitter, the apparent bearing is found to vary with time, owing to changes in the ionosphere. With continuous-wave signals there are often large bearing fluctuations within periods of the order of seconds, but it is known that much longer periods are also significant. Some data on the time scale of the most rapid variations have been given by Bain.¹ The present paper gives the results of measurements over longer periods of time, which enable the relative magnitudes of the faster and slower fluctuations to be estimated.

(2) DESCRIPTION OF EXPERIMENTS

For the main experiments a number of strong and reliable c.w. stations were chosen for regular observation over a period of three months, January–March, 1954. In addition, as pulse transmissions from Malta were available at the time, a few measurements on the first-order F-reflections from this station were included. The details of the transmitters used are given in Table 1.

Table 1

TRANSMITTERS STUDIED

Station	Frequency	Distance	Bearing
	Mc/s	km	deg
Schwarzenburg	6.165	780	128.8
Göteborg	8.754	1010	44.9
Prague	8.910	1040	91.6
Warsaw	11.534	1470	78.1
Helsinki	8.706	1850	48.9
Malta	11.210	2120	139.6
Sackville	17.82	4610	289.6
Schenectady	17.76	4990	290.2

The measurements were all made between 0930 and 1730 hours U.T., and in the case of Sackville and Schenectady they were confined to the afternoon period between 1400 and 1730 hours. Each sequence of observations on a station consisted of about 100 snap bearings, taken visually on the cathode-ray direction-finding display of an Admiralty-pattern twin-channel receiver.²

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

At the outset of the experiments the interval between snaps was made 20 sec, and the readings were taken by the method used in previous series of observations of this sort, namely by attempting to judge the instantaneous position of the cathode-ray-tube trace and setting a rotatable cursor in this position. This method was, however, soon replaced by an improved technique, in which a pushbutton switch was used to brighten the trace momentarily. A cathode-ray tube with about 5 sec afterglow was used, and the procedure for taking a bearing was simply to set the cursor on the stationary afterglow trace. The bulk of the observations were made using this technique, at a rate of either four or six per minute. In all cases the snap observations were made at regular time intervals, independently of the signal conditions. If a reading was missed owing to a fade, or to interference, the reason was noted, and also any readings taken under conditions of a fairly deep fade, as judged by the operator, were specially marked.

On each day on which the observations were made, local 'spot' instrumental calibrations were carried out with a portable oscillator on the frequencies and true bearings of the transmitters used. At the time of the experiments, access to the field in the immediate neighbourhood of the direction-finder was not available, except for the path leading to it, and so the oscillator had to be placed at the appropriate bearings outside the edge of the field, some 150–300 m from the direction-finder. For the two American stations, moreover, the reciprocal bearings had to be used, as only the eastern edge of the field was accessible. These local tests could therefore not be regarded as providing an accurate absolute calibration, but merely served to indicate any changes in instrumental balance.

(3) RESULTS

(3.1) Short-Period Variations

For each set of about 100 bearings taken in a period of 20–30 min, the mean value \bar{x} and the variance σ^2 were first calculated, using all the readings obtained. These usually contained a number of widely deviated bearings owing to the more extreme effects of wave interference. These conditions often produced small signal amplitudes, leading to comparatively large polarization errors, in addition to the wave interference errors. To eliminate such cases, therefore, the analysis was repeated after exclusion of all readings lying outside a range arbitrarily chosen as $\pm 3\sigma$, which would be expected to exclude 0.27% of the results if their distribution was normal. This process of eliminating 'wides' was repeated if necessary until all the results lay within $\bar{x} \pm 3\sigma$, \bar{x} and σ now referring to the values still remaining in the sample. As a check on the effect of fades alone, a third calculation of \bar{x} and σ^2 was made, this time excluding from the original sample only those results designated by the operator as having been taken in a fade.

Table 2 gives the overall results for each transmitter separately, and for all of them combined, for all the periods of observation. To eliminate the effect of bearing variations from period to period in combining the results, all deviations were measured from the mean of the period in which they were obtained.

Table 2
SUMMARY OF BEARING DATA

Station	Number of bearings	Percentage missed owing to fades	σ^2 (deg ²) (all results)	Percentage of wides	σ^2 (deg ²) excluding wides	Percentage of fades in bearings taken	σ^2 (deg ²) excluding fades
Schwarzenburg	3 166	0.7	13.1	2.0	8.7	2.4	11.8
Göteborg (a)	1 890	0.8	62.2	1.6	57.7	2.1	56.0
Göteborg (b)	1 757	0.5	6.4	1.6	5.0	2.0	5.4
Prague	3 434	0.9	6.0	1.8	4.3	2.4	4.8
Warsaw	2 089	1.0	3.8	2.8	2.5	2.6	3.4
Helsinki	1 679	1.0	10.7	2.1	7.6	2.9	9.3
Malta	827	5.9	2.6	1.9	2.0	2.2	2.5
Sackville (a)	2 152	2.2	35.0	2.7	21.5	2.2	33.1
Sackville (b)	1 861	2.1	13.0	2.6	7.4	2.2	12.2
Schenectady (a)	1 500	2.3	58.4	1.9	50.7	4.9	53.0
Schenectady (b)	1 400	2.4	13.3	1.9	9.1	5.0	9.8
All stations (a)	16 737	1.4	22.1	2.1	17.5	2.6	20.0
All stations (b)	16 213	1.4	8.9	2.1	6.0	2.6	7.7

It will be seen that double entries in Table 2 have been made in the cases of Göteborg, Sackville and Schenectady. For all the stations, the variances in individual periods varied considerably, over a range of about 10 : 1 (3 : 1 or 4 : 1 for standard deviations), but for the three stations just mentioned, there were some periods (two on Göteborg, three on Sackville and one on Schenectady) in which the results were quite anomalous, in respect of very large variances, many times removed from the usual range of values. As these periods weight the overall results, given in line (a) for the three stations, very heavily, a second line (b) has been added showing the results obtained when the anomalous periods are excluded. These exceptional results were obtained at normal times of observation of the stations concerned, and do not appear to have been associated with clear-cut skip conditions, although the signals were somewhat weaker than usual. On Sackville and Schenectady the transition from normal ray to skip conditions appeared to be a rather gradual effect, and there was a definite tendency for the larger variances to be associated with the weaker signals, as illustrated in Fig. 1. The figure shows the standard deviation

are rather larger, on the whole, than the values obtained³ in a similar set of measurements made previously with the same Adcock instrument, although the station which gave the largest deviations in that series (Prague, variance about 12 deg²) gave much smaller variances in the present tests. The two sets of measurements were made at the same time of year (three years apart), with very similar techniques, and the discrepant results illustrate the great difficulty of generalizing from one set of data to another, even under apparently identical conditions.

The results do not establish any systematic dependence of variance on distance or frequency. The variances appear to rise at the shorter and also at the longer distances, but to set against this must be recorded the fact that a small sample of bearings (five periods, 471 observations) was also taken on Schenectady on a lower frequency, 15.32 Mc/s, and these yielded a consistently small variance of about 2.5 deg².

Sets of observations were frequently obtained on Sackville 17.82 Mc/s and Schenectady 17.76 Mc/s, in periods within an hour or two of each other, on the same days. These stations, operating on very nearly the same frequency and with their true bearings less than a degree apart, might have been expected to show some similarity in behaviour, but it was found that there was no correlation between the variations of either mean bearing or variance obtained on the two stations. Information has been obtained concerning the transmitting aerials used at Sackville and Schenectady during the period of the observations, by courtesy of the Canadian Broadcasting Corporation, and the General Electric Co., respectively. The transmissions were beamed towards Europe in both cases, and the polar diagrams in both horizontal and vertical planes were very similar. Each had a major lobe about 35° wide at 6 dB down from the maximum. The Schenectady beam was unchanged throughout the tests; at Sackville two orientations, 10° apart, were used on different occasions, but the direction of shoot was always within 10° of the great-circle bearing of Slough. Thus the beam changes would hardly be expected to affect the d.f. results, and no recognizable effects were in fact observed. The diverse results obtained on the Sackville and Schenectady transmissions cannot therefore be ascribed to different transmitter aerial characteristics.

The analysis of the effects of excluding wides and fades from the results shows that only about 2% of readings fall into either category. The exclusion of wides results in a variance reduction of about 30% on the average, but the exclusion of fades, although these were slightly more numerous than the wides, brought about a smaller reduction, of about 10% only. Thus although the

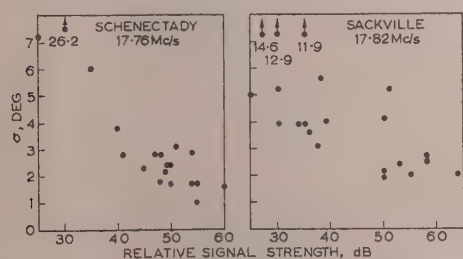


Fig. 1.—Variation of standard deviation with signal strength.

for a period plotted against the approximate mean signal level in decibels during the period, relative to an arbitrary zero. A similar effect is suggested weakly in the Warsaw and Helsinki data, but is not apparent in the results from the other stations. It thus appears to become more pronounced with increasing distance, and it may be noted that Sackville and Schenectady are the only ones for which the transmission path would require more than a single hop.

With regard to the general magnitude of the variances, the results illustrate the very large bearing variations which occur when truly snap readings are taken on c.w. signals. The variances

results confirm that greater errors tend to be produced under fading conditions, the effect is only slight, and moreover the fades by no means account for all the wide bearings obtained.

The variance obtained from the 1F pulse measurements on Malta is, with the exception of the 15 Mc/s Schenectady results mentioned above, smaller than any of the c.w. variances. This was to be expected since intermode interference was, of course, absent in the pulse observations. It is of interest to compare the Adcock variance with the corresponding figure obtained from simultaneous observations made on the same transmissions using a wide-aperture phase-comparison d.f. system.⁴ Using only periods when such joint observations were made, the Adcock results gave a variance of 1.6 deg^2 and the corresponding result for the wide-aperture system was 0.17 deg^2 . It is thus clear that the deviations in the true direction of arrival of the signal are much magnified in the Adcock results. A marked difference in variance, often larger than the above and in some cases up to about 5 deg^2 , has also been obtained in previous joint observations with the Adcock and the wide-aperture system, on pulse transmissions from Sterling, U.S.A.,⁵ from Oslo and from Inverness. This extra variance must be ascribed to variable errors associated with the direction-finder itself and the surrounding site.

(3.2) Medium-Period Variations

Next will be considered the variations of a 5 min mean bearing, corresponding to what has often been referred to as the slowly varying component due to lateral deviation of the ionospheric waves. The analysis of the present results has been carried out by first calculating the mean bearing for each consecutive five minutes of observation, involving 15–30 snap readings, and the deviation of this mean from the mean of the whole period of 20–40 min. The mean-square value of this deviation over all the data was then obtained for each transmitter separately. In this analysis no wides or fades were rejected, but as before the calculations were repeated for Göteborg, Sackville and Schenectady, after omitting the anomalous periods. Table 3 shows these

nominal interval between observations, owing to occasional bearings being missed.

These results on time averaging are intimately connected with the question of the correlation existing between individual bearings at various time intervals. Thus if ρ_j is the correlation coefficient between readings j intervals apart, it is easily shown that

$$\frac{\sigma_n^2}{\sigma_1^2} = \frac{1}{n^2} \left[n + 2 \sum_{j=1}^{n-1} (n-j) \rho_j \right] \quad (1)$$

Table 3 shows that σ_n^2/σ_1^2 is always greater than $1/n$, which corresponds to the case of zero correlation between individual readings.

If a correlation function of the form $\rho(\tau) = e^{-\tau/\tau_0}$ is assumed, so that $\rho_j = \rho_1^j$, eqn. (1) becomes

$$\frac{\sigma_n^2}{\sigma_1^2} = \frac{n(1 - \rho_1^2) - 2\rho_1(1 - \rho_1^n)}{n^2(1 - \rho_1)^2} \quad (2)$$

Values of ρ_1 , the correlation between consecutive bearings, can be calculated from the latter equation using the measured values of σ_n^2 and σ_1^2 given in Table 3, and are found to range from zero up to 0.7, with an average of about 0.3. This would imply an effective time-constant τ_0 of the order of 8 sec for the bearing fluctuations. However, a closer analysis of the results shows that the autocorrelation is not of the simple exponential form assumed above, when observations over a period of twenty minutes are considered. Actual values of the autocorrelation existing at intervals of 10 sec, 1 min and 2 min have been computed from the bearings taken at 10 sec intervals, and the mean results for each station are shown in Table 4. It is seen that for the c.w. stations the correlation is practically zero on the average at 2 min, but it is consistently positive, though very small, at 1 min. The Malta pulse results showed greater correlation at 1 and 2 min; it was found to fall to zero at between 5 and 10 min, as previously found in pulse measurements at shorter distances.⁶ In Table 4 the ratio (σ_n^2/σ_1^2) of the 5 min-mean variance to that

Table 3

REDUCTION OF VARIANCE BY AVERAGING OVER 5 MINUTES

Station	20 sec intervals			15 sec intervals			10 sec intervals		
	n	σ_1^2	σ_n^2	n	σ_1^2	σ_n^2	n	σ_1^2	σ_n^2
Schwarzenburg	—	—	—	19	9.4	1.50	28	14.0	0.88
Göteborg (a)	—	—	—	14	7.4	1.39	18	82.8	16.5
Göteborg (b)	—	—	—	14	7.4	1.39	18	6.1	1.46
Prague	14	6.2	0.92	18	4.5	0.36	28	6.0	0.47
Warsaw	13	3.1	0.74	19	2.8	0.35	27	4.2	0.19
Helsinki	13	6.0	0.97	18	11.0	1.59	26	11.9	1.09
Malta	11	5.2	0.89	17	2.0	0.51	—	—	—
Sackville (a)	13	15.2	1.42	17	45.3	3.25	26	37.3	2.57
Sackville (b)	13	15.2	1.42	17	8.6	0.51	26	13.5	1.51
Schenectady (a)	14	5.6	0.72	19	9.1	0.78	26	122	8.40
Schenectady (b)	14	5.6	0.72	19	9.1	0.78	26	21.5	3.13
All stations (a)	13	7.9	0.97	18	10.2	1.18	26	29.6	4.16
All stations (b)	13	7.9	0.97	18	7.0	0.93	26	9.8	1.00

5 min mean variances σ_n^2 for each station, and for all stations together, in three groups according to the rate at which observations were taken, and also the variance σ_1^2 of the individual readings taken in the appropriate rate group. Also shown are the average number of readings n actually included in each 5 min period; this number is less than would be expected from the

of the individual values, for the results taken at 10 sec intervals, is also shown for each station (these ratios are obtained from Table 3); it is seen that the ratio is higher, as expected, in the cases of the larger values of correlation. At the foot of the Table the mean values of the correlations for the c.w. stations are given, and also the corresponding mean values of σ_n^2/σ_1^2 .

Table 4

CORRELATION BETWEEN BEARINGS TAKEN AT VARIOUS INTERVALS

Station	Correlation coefficient			σ_b^2/σ_f^2
	10 sec	1 min	2 min	
Schwarzenburg	0.13	0.06	-0.04	0.06
Göteborg (a)	0.27	0.08	-0.08	0.20
Göteborg (b)	0.28	0.05	-0.08	0.24
Prague	0.12	0.03	0.01	0.08
Warsaw	0.10	0.01	-0.03	0.04
Helsinki	0.13	0.15	0.08	0.09
Sackville (a)	0.24	0.06	0.05	0.07
Sackville (b)	0.25	0.07	0.09	0.11
Schenectady (a)	0.10	0.06	0.06	0.07
Schenectady (b)	0.09	0.07	0.10	0.14
Malta	—	0.14	0.18	—
Mean of c.w. stations (a) ..	0.16	0.06	0.01	0.09
Mean of c.w. stations (b) ..	0.16	0.06	0.02	0.11

The correlation data may be used to calculate the degree of variance reduction obtainable by continuous averaging over various time intervals. For continuous averaging over time τ , eqn. (1) becomes

$$\frac{\sigma_f^2}{\sigma_b^2} = 2 \int_0^1 (1-u) \rho(u\tau) du = F(\tau), \text{ say} \quad (3)$$

$F(\tau)$ may be called the variance-reduction function. In order to calculate its value for any time τ it is necessary to specify the correlation function for all intervals from 0 to τ . The measured values of ρ given at the foot of Table 4 have therefore been used to construct, tentatively, a curve giving ρ as a continuous function of τ , and this is shown in Fig. 2. No information concerning the

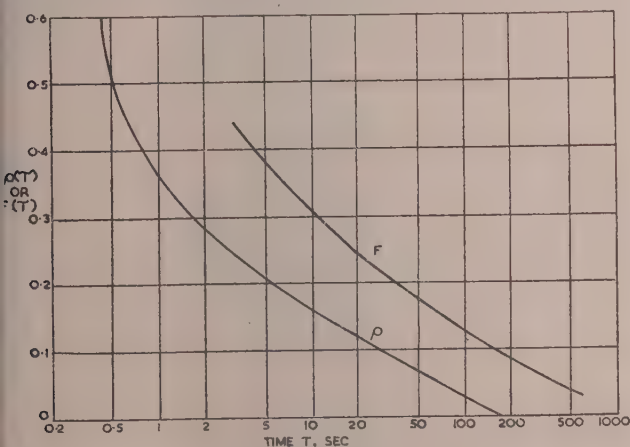


Fig. 2.—Auto-correlation and variance-reduction functions.

value of ρ for τ less than 10 sec could be obtained from the present results; in this region the curve has been drawn by assuming the correlation to fall exponentially from unity, for values of τ up to about $\frac{1}{2}$ sec, at a rate corresponding to a time-constant of $\frac{1}{2}$ sec, as given by Bain's measurements,¹ and these values have then been joined up smoothly with those deduced from the present experiments.

Using values of ρ as given by this curve, the variance-reduction function has been obtained from eqn. (3) by numerical integration. The resulting curve is also shown in Fig. 2. Regarding the

use of the curves in this Figure it must be emphasized that they represent only average results, and considerable deviations from these may occur on particular occasions, depending on the relative magnitude of the faster and slower components of the bearing fluctuations at the time. Also, the part of the curves for small values of τ must be regarded as rather tentative.

The value of $F(\tau)$ as shown by the curve in Fig. 2 is, for any particular averaging period τ , greater than the value implied by the results of Bain's observations of the rapid fluctuations. These indicated that the variance would be reduced by a factor of 10 in 12 sec, whereas the present results give a smaller, but still useful, factor of $3\frac{1}{2}$ in this time. This is due to the fact that in Bain's measurements the total time of observations was less than a minute, so that no account could be taken of fluctuations with periods longer than this. The present results show that these slower variations, which are already well known to be of major importance when the rapid variations due to wave interference are removed, are still significant when dealing with snap observations where the variance of the individual bearings is quite large.

It should be noted that the 5 min-mean variations, as recorded on the Malta pulse transmissions, are definitely larger than can be attributed to true propagational deviations; as measured with the wide-aperture system⁷ these amount to only 0.1 deg^2 . Other examples of the increased variance of 5 min means on pulse transmissions, as compared with wide-aperture measurements, have been obtained in the Oslo, Sterling, and Inverness results previously mentioned. An extra variance of 1.2 deg^2 is always added to the wide-aperture values, which seldom exceed 0.5 deg^2 . The extra variance must again be ascribed to instrumental and site errors.

(3.3) Long-Period Variations

Long-period variations are the variations in bearing occurring over intervals of several hours on the same day, or from day to day. To investigate their magnitude the mean bearing for each group of about 100 observations on each transmitter was taken, and the variance of these mean values calculated for each station separately. The calculation was repeated after the application of corrections based on the local calibration made on the same day as the bearing observations. The results are shown in Table 5, together with the number of periods on which they are based. The Göteborg and Schwarzenburg results have been excluded as they are considered unreliable owing to receiver line-up difficulties involved in observations on these transmitters; this would not appreciably affect the results previously described on the shorter period variations. As before, the Sackville and Schenectady analyses are given, both including and excluding the anomalous sets of results. Of the six anomalous periods on these two stations and Göteborg, four gave mean bearings several degrees different from the overall average for the transmitter concerned, and in each case these deviations represented an apparent shift of the transmitter towards the south. The Table also shows the mean bearing errors with respect to the correct bearings, for all the periods of observation, both with and without the application of local calibration corrections. These show quite large errors, but as previously mentioned, the local calibrations themselves were probably subject to considerable errors.

Table 5 shows that, in five cases out of six, the variance of the period mean is reduced by application of the local calibration correction when the anomalous results are ignored; if they are included the variance is decreased in three cases and increased in the other three. The effect of these corrections is not large in any case, as the actual variations in the local calibrations during the period of the experiments were generally small, with a

Table 5
LONG-PERIOD BEARING ERRORS

Station	Number of periods	Variance of period mean		Sampling variance	Mean error	
		Without local corrections	With local corrections		Without local corrections	With local corrections
		deg ²	deg ²	deg ²	deg	deg
Prague	33	1.8	1.0	0.13	2.3	1.7
Warsaw	21	0.9	1.6	0.07	0.6	3.5
Helsinki	19	2.3	2.0	0.33	1.5	-0.5
Malta	8	2.0	0.6	0.09	4.2	4.2
Sackville (a)	22	2.5	2.6	0.75	2.2	-2.7
Sackville (b)	19	1.7	1.4	0.39	2.6	-2.3
Schenectady (a)	15	20.8	26.1	1.5	3.3	-1.4
Schenectady (b)	14	1.6	1.3	0.40	4.5	-0.1
All stations (a)	118	4.3	5.1	0.44	1.5	0.6
All stations (b)	114	1.7	1.7	0.23	1.7	1.0

variance of about 1 deg². It can be seen that whether local calibrations are applied or not, there are substantial variations in the period-mean bearings. Some of these variations are due to sampling fluctuations, and an estimate of the variance attributable to this is included in the last column of the Table. This was calculated from the number of observations and their variance within individual periods, due allowance being made for the correlation found within a period, and represents the expected variance of the period means on the hypothesis that there was no real variation from period to period. In all cases the actual variances are much greater than those expected on this hypothesis, so it may be concluded that the period-to-period variations are real. Attention does not appear to have been directed to these variations in the past, although it seems that they are of quite a serious magnitude. From the results described above it appears unlikely that the variations represent real bearing changes of the incident radiation, which, as measured on a wide-aperture direction-finder using pulse transmissions,⁷ are much smaller. To obtain more information on the effect some further measurements have been carried out as described below.

(4) FURTHER INVESTIGATION OF THE LONG-PERIOD VARIATIONS

It was first verified that the variations shown in Table 5 were not due to errors in the twin-receiver line-up by taking a series of measurements on a single transmitting station in which a half-hour period of observation using the visual display was immediately followed by a similar period using a manually operated goniometer (aural null) system connected to the same Adcock

aerials. The latter equipment, using only a single receiver, was, of course, free from the line-up necessary with the twin-channel visual system. It was found that the period-mean bearings in adjacent periods on the two receiving systems were well correlated, while the period-to-period variations on each were similar in magnitude to those previously obtained. It was thus clear that line-up errors were not responsible for these variations.

Next, in order to examine the magnitude of the variations on a given day, as compared with those occurring from day to day, observations were made on three separate days on transmissions from Prague (frequency, 8.91 Mc/s) for a continuous period of seven hours or more on each day. The bearings were taken visually on the cathode-ray direction-finding display, at 10 sec intervals. Calibrations with a local oscillator were made at two-hourly intervals, and receiver line-up checks every half-hour. Small errors in receiver line-up were, however, relatively unimportant in their effect on the indicated bearing, since this was near 90°. The local calibrations did not vary significantly during any one day, but differed appreciably on the different days.

From the individual snap bearings, mean values were calculated in half-hour periods, and the variations of the period means are summarized in Table 6. U and C, respectively, denote the results uncorrected and corrected by the application of the calibration.

The last column shows that sampling fluctuations account for only a small proportion of the observed period-mean variations. The total variance of 0.8 deg² is slightly less than that obtained (1.0 deg²) for this transmitting station in the previous experiments, which covered a much longer period. The results indicate

Table 6
PERIOD-MEAN BEARING VARIATIONS ON PRAGUE

Date, 1955	Number of ½-h periods	Mean bearing error		Range of period means		Variance of period means		Sampling variance of period means
		U	C	U	C	U	C	
		deg	deg	deg	deg	deg ²	deg ²	deg ²
19th January	17	+1.6	+1.2	2.4	2.4	0.4	0.4	0.03
15th February	19	+1.3	-0.1	2.9	2.9	0.4	0.4	0.06
25th March	14	+4.2	+1.3	2.3	2.3	0.5	0.5	0.10
All days	50	+2.2	+0.7	5.1	2.9	2.0	0.8	0.06

at variations from period to period on the same day, and variations from day to day, are of about the same magnitude.

For the final set of tests a small rotating spaced-loop direction-finder⁸ (aerial spacing, 3 m) was installed in a hut about 100 m from the Adcock, and bearings were taken on six distant stations. Each period of observation on a station was half an hour, and in any period the same station was observed at each direction-finder. All the bearings were taken aurally at half-minute intervals, and each bearing was an average over a few seconds, being the result of several swings of the goniometer in the case of the Adcock, or the aerial system in the case of the spaced-loop direction-finder. The tests were made on 11 days from 19th July to 8th August, 1955, and not more than two periods of observation were made on any one station in any day.

On each day, local calibrations of the Adcock were made on the frequencies and bearings of the stations observed on that day, so that allowance could be made for day-to-day changes in the instrumental error, or very local site error. Such changes did not exceed $1\frac{1}{2}^\circ$ for any frequency and bearing during the period of the tests, and the Adcock results were analysed both with and without the application of corrections based on these local calibrations. No calibrations of the spaced-loop direction-finder were made.

A significant feature of the results is the absence of correlation between the period-mean variations on the two direction-finders. This confirms the view, previously suggested, that the variations are not real bearing changes of the received signals. Their actual cause is considered further below.

(5) SUMMARY AND DISCUSSION OF RESULTS

The results described have shown the existence of bearing fluctuations on an Adcock direction-finder having characteristics broadly as follows:

(a) Variations within a period of 20 to 30 min having a variance of about 10 deg^2 for snap bearings. The great variability of this figure from time to time must, however, be stressed.

(b) Variations of 5 min means within this period with a variance of about 1 deg^2 . These variations are accounted for by a correlation function for the bearing fluctuations which falls much less rapidly, at intervals greater than a few seconds, than the function previously found by Bain for shorter intervals; it is still appreciably positive at intervals up to at least a minute.

(c) Variations from period to period with a variance of about 1 deg^2 .

The first of the above three types of variation can be ascribed mainly to wave-interference effects between separate modes of propagation, each of which is subject to small bearing deviations.

Table 7

COMPARISON OF PERIOD-MEAN BEARING VARIATIONS ON ADCOCK AND SPACED-LOOP SYSTEMS

Station	Frequency	Distance	Bearing E of N	Number of periods	Mean bearing error			Range of period means			Variance of period means			Variance ratio	
					A(U)	A(C)	S	A(U)	A(C)	S	A(U)	A(C)	S	A(U)/S	A(C)/S
	Mc/s	km	deg		deg	deg	deg	deg	deg	deg	deg ²	deg ²	deg ²		
Allouis ..	6.200	520	156.9	13	-2.0	-0.6	-1.0	3.5	3.3	3.3	0.9	1.0	1.2	0.8	0.8
Free Hungary	5.970	720	106	18	-3	-3	0	3.0	3.0	1.2	0.6	0.7	0.1	6	7
Schwarzenburg	6.165	780	128.8	18	+0.7	+2.8	+0.4	4.8	5.0	2.2	1.3	1.5	0.3	4	5
Prague ..	10.505	1040	91.6	15	-0.3	+0.6	+0.3	2.6	3.1	1.7	0.6	0.8	0.3	2	3
Jeloy ..	9.980	1070	34.3	9	+0.5	-2.4	-0.5	4.0	4.3	2.8	1.2	1.4	0.8	2	2
Sackville ..	15.09	4610	289.6	8	+6.8	-1.8	-0.4	2.6	1.9	1.1	0.5	0.4	0.1	5	4
All stations ..				81	-0.2	-0.5	-0.1	4.9	5.4	3.6	0.9	1.0	0.4	2	2

* This station is near Stuttgart but its position is not known exactly.

Table 7 summarizes the results. For each transmitting station, the variances of the period means as observed on the Adcock (A) and the spaced loop (S) are given, and also the ratio of these variances, both before and after applying calibration corrections to the Adcock results [A(U) and A(C), respectively]. The last line of the Table gives the figures obtained by pooling the results on all the stations, the period-mean variations for each station being referred to its own grand mean bearing.

It is seen that the Adcock variances are again of the order of deg^2 . They are actually increased slightly by application of the calibration corrections, and so is the overall mean bearing error, though the maximum error for any station is reduced from 6.8° to 2.8° . The spaced-loop variations from period to period are appreciably smaller than those on the Adcock, for all stations except Allouis. The overall value of 0.44 deg^2 for the spaced loop is significantly less, at the 1% level, than either of the Adcock values. As before, only a small proportion of the period-to-period variance can be attributed to sampling errors. In the present case the sampling variance did not exceed about 0.05 deg^2 , as the variance of the individual bearings within a period was only up to about 3 deg^2 owing to the time-averaging carried out in taking the bearings.

The residual variance of 2.6 deg^2 obtained on pulse transmissions, where the relatively large multimode interference effects are eliminated, must be attributed mainly to site, instrumental (including polarization error) and possibly observational errors, since the true bearing deviations of the incident signal, as measured on a wide-aperture system, have a much smaller variance still.

The 5 min-mean variations, although much smaller than the more rapid fluctuations, are also larger than can be accounted for by true ionospheric deviations. The extra variance must again be put down to site and instrumental errors, which are sufficiently systematic not to average out completely in 5 min.

The origin of the long-period variations is still in some doubt. It has been shown that they are not real changes in the bearing of the downcoming waves from the ionosphere; it is also fairly certain that they are not due to purely instrumental errors, since the Adcock variations are not accounted for by local calibration corrections, and the spaced-loop variations are larger than could be attributed to variable instrumental errors.

The most likely explanation of the long-period effects appears to be that they represent variations from period to period of the distant-site error, caused by changes in the true direction of

arrival of the incident waves. It is known⁹ that true bearing changes of an incident wave can be magnified by distant-site effects. Further, changes in the angle of elevation would similarly be expected to produce apparent bearing errors, even though the actual bearing remained correct, and previous pulse experiments have shown that, in general, changes in elevation are greater than the corresponding bearing variations. Theoretical analysis of the relative magnitudes of site errors on different direction-finders, based on averaging the effect of a single re-radiator revolved around the system, indicates¹⁰ that the site-error variance for a narrow-aperture spaced-loop direction-finder should be less than that for an Adcock by a factor of 4, owing to the different polar diagrams of the two systems. It may be shown that, if a model more closely related to the practical situation is assumed, namely with a large number of re-radiators randomly distributed around the direction-finder, the same ratio of four is obtained. The value of this factor given by the present results is only about two, but statistically they are not inconsistent with the hypothesis that the true value is four.

It is of interest to compare the bearing variances obtained in the present experiments with previously published estimates. Smith and Hopkins suggest¹¹ a total variance about the correct bearing of $9-10 \text{ deg}^2$ for snap bearings on c.w. transmissions and $2.5-5 \text{ deg}^2$ on pulse transmissions, first-order E-echo. The present results for c.w. and pulse, first-order F echo, are quite consistent with these figures. The results for 5 min means are also in general agreement, although a strict comparison is not possible since estimates of absolute accuracy have not been attempted in these experiments.

(6) ACKNOWLEDGMENTS

The author acknowledges the assistance of Messrs. A. Ambery, L. A. Bonvini, W. A. Etheridge and P. T. A. Spurle in the observational work and numerical analysis of the results.

The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(7) REFERENCES

- (1) BAIN, W. C.: 'On the Rapidity of Fluctuations in Continuous-Wave Radio Bearings at High Frequencies', *Proceedings I.E.E.*, Paper No. 1715 R, October, 1954 (102 B, p. 541).
- (2) DE WALDEN, S., ROCKE, A. F. L., BARRETT, J. O. G., and PITTS, W. J.: 'The Development of a High-Frequency Cathode-Ray Direction-Finder for Naval Use', *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 823.
- (3) HOPKINS, H. G., and BRAMLEY, E. N.: 'Some Practical Measurements of the Relative Performances of a Cyclical Phase-Comparison Type of Direction-Finder and a U-Adcock Instrument', *Proceedings I.E.E.*, Paper No. 1544 R, September, 1953 (100, Part III, p. 263).
- (4) ROSS, W., BRAMLEY, E. N., and ASHWELL, G. E.: 'A Phase-Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves', *ibid.*, Paper No. 1134 R, July, 1951 (98, Part III, p. 294).
- (5) McCUE, C. G.: 'H.F. Direction Finding: Observations on a Transmitter of Adjustable Beam-Direction', *Wireless Engineer*, 1955, **32**, p. 79.
- (6) BRAMLEY, E. N.: 'Direction-Finding Studies of Large-Scale Ionospheric Irregularities', *Proceedings of the Royal Society, A*, 1953, **220**, p. 39.
- (7) BRAMLEY, E. N.: 'Directional Observations on High-Frequency Transmissions over 2100 km', *Proceedings I.E.E.*, Paper No. 2011 R, May, 1956 (103 B, p. 295).
- (8) ROSS, W.: 'The Development and Study of a Practical Spaced-Loop Radio Direction-Finder for High Frequencies', *Journal I.E.E.*, 1947, **94**, Part III, p. 99.
- (9) ROSS, W.: 'Site and Path Errors in Short-Wave Direction-Finding', *ibid.*, p. 108.
- (10) HOPKINS, H. G., and HORNER, F.: 'Direction-Finding Site Errors at Very High Frequencies', *Proceedings I.E.E.*, Paper No. 773 R, January, 1949 (96, Part III, p. 321).
- (11) SMITH, S. B., and HOPKINS, H. G.: 'H.F. Direction Finding: Comparison of C.W. and Pulsed Transmissions', *Wireless Engineer*, 1954, **31**, p. 11.

THE CONTINUOUS DELAY-LINE SYNTHESIZER AS A SYSTEM ANALOGUE

By J. H. WESTCOTT, B.Sc.(Eng.), Ph.D., Associate Member.

(The paper was first received 16th September, 1955, and in revised form 26th January, 1956. It was published as an INSTITUTION MONOGRAPH in May, 1956).

SUMMARY

The use of a continuous delay-line synthesizer as a means for simulating transfer functions of linear systems by a very simple setting procedure is described. An analysis of the accuracy of this device as system analogue is given, and the degree of approximation in frequency response is calculated for a number of transfer functions typical of servo systems. It is concluded that a delay-line of 26 sections will give a frequency response with not more than 5% error in the working band of frequencies for all normal servo systems.

LIST OF PRINCIPAL SYMBOLS

- $a_1, a_2 \dots a_k$ = Proportions of voltages tapped off from the delay-line.
- $H(j\omega)$ = True frequency response of the system to be imitated.
- $H(t)$ = True step-function response of the system to be imitated.
- $\bar{H}(j\omega)$ = Approximation to the frequency response of the system provided by the delay-line synthesizer.
- $\bar{H}(t)$ = Approximation to the step response of the system provided by the delay-line synthesizer.
- T = Longest delay provided by the delay line.
- τ = Interval between sampling points.
- N = Total number of tapping points along the delay line.
- $\omega_0 = 2\pi/\tau$ = Period of the harmonics in the imitated frequency response (see Fig. 2).

(1) INTRODUCTION

It has been shown^{1,2} that a sectioned delay-line may be used to realize the indicial response (response to a step-function input) of a system. The required response is synthesized directly in terms of incremental contributions of the same form delayed from each other by equal time intervals and adjusted by independent settings. The resulting response then approximates the indicial response of the system in a stepped fashion or by means of simple curves. It may be expected that the delay line so adjusted will exhibit a frequency response approximating that of the original system. The paper discusses the order of accuracy of this last approximation if a continuous delay-line is used.

(2) THE CONTINUOUS DELAY-LINE ANALOGUE

(2.1) Method of Application

The method of realizing the system transfer function by means of a linear continuous delay-line is shown schematically in Fig. 1. The input signal proceeds without distortion along the delay line and is tapped off at equispaced time intervals, τ , in proportion to coefficients a_k (which are adjustable). All such delayed contributory signals are then summed to give an output resembling the response of the simulated system. When the input is a unit step-function, as shown in the Figure, the output is composed

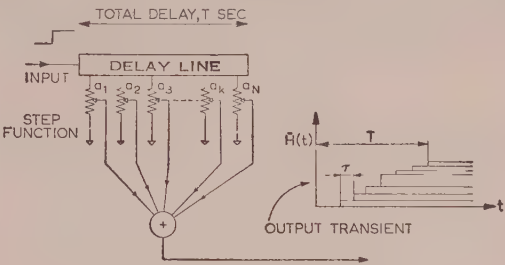


Fig. 1.—The delay-line analogue.

of a number of superimposed and delayed step-functions in which the k th tap along the delay line contributes a step of magnitude a_k commencing after a time $k\tau$ sec. The total delay of the line is T sec with a tap every τ sec; giving a total number of taps $N = T/\tau$. However, if the input is continuous, the output is also continuous.

(2.2) Sources of Error in the Analogue Representation

There are two components of error in representing the system transfer function by means of the analogue, namely

- (a) Inaccuracies due to the finite delay-section interval τ ; the smaller τ can be made, the closer the approximation.
- (b) Inaccuracies arising from the finite length of the total delay time, T , of the delay line; the longer T can be made, the closer the approximation.

The sampling interval τ gives rise to a response whose Fourier spectrum is periodic in an interval of frequency $1/\tau$, as shown in Fig. 2.

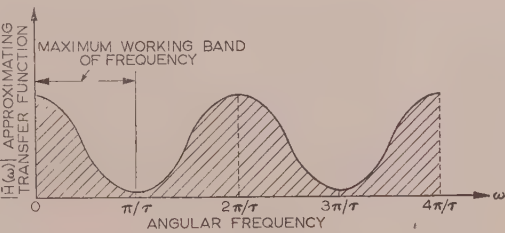


Fig. 2.—Frequency response of the delay-line analogue.

It is clear that, for an accurate representation of a response containing frequencies up to f_s , the sampling interval τ may not be greater than $\tau = 1/2f_s$.

The finite duration of the total delay time of the line, T , gives rise to error due to the components a_k that are neglected when k would be greater than N , the total number of delay intervals. The indicial response of a linear system has an infinite duration, but since the delay line has a total delay T , it is necessary to make a decision after T sec on how the indicial response is to be terminated in the approximation to it. Various choices are possible, and it is clear that they will affect the degree of error that results. The simplest method of termination is to make the

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Westcott is in the Electrical Engineering Department, Imperial College of Science and Technology, University of London.

magnitude of the imitating response for all times greater than T equal to the magnitude of the system indicial response at T ; in which case the effect of the neglected coefficients a_k for $k > N$ gives directly the error involved. However, it may be noted that this is not necessarily the best choice.

(3) A DETAILED ANALYSIS

The output of the delay line is the sum of the outputs of simple delaying systems provided by its tapping points. The frequency response of the delay line, $\bar{H}(j\omega)$, is therefore equal to the sum of the frequency responses of these simple delaying systems:

$$\bar{H}(j\omega) = \sum_{k=1}^N a_k \varepsilon^{-j\omega k\tau} \quad (1)$$

If the proportions of the outputs from the delay-line tapping points are adjusted to give an approximation as shown in Fig. 3,

$$a_k = H(k\tau) - H(k\tau - \tau) \quad (2)$$

where $H(t)$ is the indicial response to be represented.

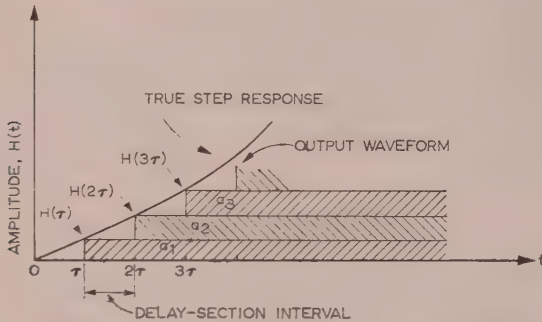


Fig. 3.—Output waveform from delay-line analogue.

Introducing this into eqn. (1) and using $H(0) = 0$ gives

$$\bar{H}(j\omega) = (1 - \varepsilon^{-j\omega\tau}) \sum_{k=0}^N H(k\tau) \varepsilon^{-j\omega k\tau} + H(N\tau) \varepsilon^{-j\omega(N+1)\tau} \quad (3)$$

In expression (3) it is assumed that the final response of the delay-line synthesizer never reaches unit displacement, since further terms a_k according to eqn. (2) for $k > N$ required to achieve this are omitted.

In many systems, however, the zero-frequency ordinate is important and for this to be correct in the representation it may be necessary for the final value of the output of the delay-line synthesizer to be unity. This can be achieved by replacing $H(N\tau)$ in eqn. (2) by the unit displacement, so that

$$\bar{H}(j\omega) = (1 - \varepsilon^{-j\omega\tau}) \sum_{k=0}^N H(k\tau) \varepsilon^{-j\omega k\tau} + \varepsilon^{-j\omega(N+1)\tau} \quad (4)$$

These alternative terminations represent extremes; there are a variety of possibilities between the two.

The error in the approximating transfer functions $\bar{H}(j\omega)$ according to eqns. (3) and (4) is the total error arising from the truncation of $H(t)$ and from the stepped response. The part of the error arising from this latter cause may be assessed separately. The error due to stepping with a delay-line having an infinite number of delay sections of τ sec may be evaluated as follows, using expression (3) or (4) with $N = \infty$:

$$\bar{H}(j\omega) = \varepsilon^{-\frac{1}{2}j\omega\tau} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} j\omega \sum_{k=-\infty}^{\infty} H(k\tau) \varepsilon^{-j\omega k\tau} \quad (5)$$

which can be written, using the Poisson summation rule,⁴

$$\bar{H}(j\omega) = \varepsilon^{-\frac{1}{2}j\omega\tau} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} j\omega \sum_{m=-\infty}^{\infty} \frac{H(j\omega + jm\omega_0)}{j(\omega + m\omega_0)} \quad (6)$$

where $\omega_0 = 2\pi/\tau$.

As shown in Fig. 2, this transfer function is periodic in $\omega_0 = 2\pi/\tau$. From eqn. (6) it may be concluded that the relative error

$$\Delta\tau = \left[\frac{\bar{H}(j\omega) - H(j\omega)}{H(j\omega)} \right]$$

may be divided into two parts, namely

(a) A contribution

$$\frac{j\omega}{H(j\omega)} \sum_{m=-\infty}^{\infty} \frac{H(j\omega + jm\omega_0)}{j(\omega + m\omega_0)}$$

arising from the overlapping of the fundamental term ($m = 0$) in the sum of eqn. (6) by the different harmonics ($m \neq 0$).

(b) A contribution

$$\varepsilon^{-\frac{1}{2}j\omega\tau} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} - 1$$

arising from the deviation of the weighting factor $\varepsilon^{-\frac{1}{2}j\omega\tau} (\sin \frac{1}{2}\omega\tau) / (\frac{1}{2}\omega\tau)$ from unity.

It is already clear from Fig. 2 that for an accurate representation τ must be chosen small enough for the important frequencies of the imitated transfer function to lie well within the semi-period frequency range $0 \leq \omega \leq \pi/\tau$, their maximum being about $\pi/2\tau$, say. The transfer functions of most imitated systems of practical interest fall off very rapidly beyond this maximum frequency so that part (a) of the relative error is negligible as compared with part (b)* and we can write

$$|\Delta\tau| \approx \left| 1 - \varepsilon^{-\frac{1}{2}j\omega\tau} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \right| \quad (7)$$

(3.1) Mid-Interval Stepped Approximation

It will be noted that the factor $\varepsilon^{-\frac{1}{2}j\omega\tau}$ in eqn. (7) is a serious contributor to error, and it is clear that a much greater accuracy may be achieved by disposing of it by simply shifting the indicial response of the delay-line synthesizer half a delay interval to the left, as illustrated in Fig. 4.

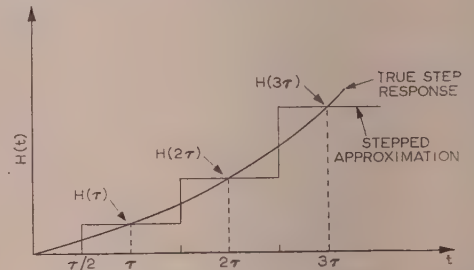


Fig. 4.—Output waveform for mid-interval stepped approximation.

The first delay section will then need to have a delay of only $\frac{1}{2}\tau$ instead of τ like the others. One can, however, arrive at a similar result by an approximation of the type shown in Fig. 5, with all delay sections the same, a negligible difference arising from

$$\sum_{k=-\infty}^{\infty} H(k\tau + \frac{1}{2}\tau) \varepsilon^{-j\omega(k\tau + \frac{1}{2}\tau)}$$

* This is still true for the much more accurate approximations to be considered later.

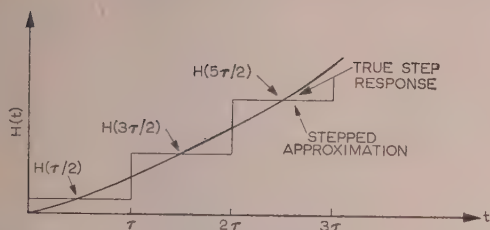


Fig. 5.—Output waveform for equal delay-interval stepped approximation.

making the place of the sum in eqn. (5) and therefore

$$\sum_{m=-\infty}^{\infty} e^{jm\tau} \frac{H(j\omega + jm\omega_0)}{j(\omega + m\omega_0)}$$

making the place of the sum in eqn. (6).

We shall refer to both of these approximations as mid-interval stepped approximations. The modulus of their relative error due to the magnitude of τ is

$$|\Delta\tau| \simeq 1 - \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \quad \dots \quad (8)$$

It will be noticed in eqn. (8) that these approximations do not give rise to any phase error.

(3.2) Straight-Line Approximation

The stepped type of waveform which may form the output with a mid-interval approximation to a system response has one unfortunate practical limitation when used in conjunction with stabilizing networks having derivative action. The steps on differentiation become impulses which are liable to suffer severe clipping in any subsequent amplifiers of limited linear range, resulting in considerable mutilation of the response. This difficulty can be avoided by making use of a straight-line approximation to the system step response instead of a stepped one, so that even after differentiation a finite step is the most severe signal that occurs. In this Section the accuracy of such an approximation is investigated.

The delay line itself can work only in a currency of step waves in response to a step-wave input, and in order to realize a straight-line approximation to the indicial response it is first necessary to establish a function resembling the impulse response of the system in the delay line when the input is a step function. This impulse response will be generated in stepped form and is then passed through an integrating circuit, emerging as a straight-line approximation to the system indicial response $H(t)$. The settings of the delay line are adjusted so that the final straight-line response assumes correct values at intervals τ . The response is thus of the form shown in Fig. 6. The first sector can be realized only if provided the first tapping point has no delay, which is henceforth assumed to be the case.

It will be observed that the shaded area is equal to that of the isosceles triangle below it (triangles on the same base and between the same parallels) and that the whole response may be represented by a set of such isosceles triangles centred at the points $k\tau$ and having the respective ordinates of the response. The frequency response of a system whose output to a step-function input is the triangle shown is as follows:

$$\bar{H}(j\omega)|_{3\tau} = H(3\tau)j\omega\tau \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \right)^2 e^{-j\omega 3\tau} \quad \dots \quad (9)$$

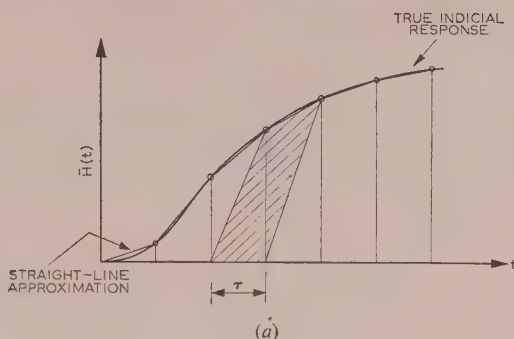


Fig. 6.—Analysis of output waveform for straight-line approximation.

Thus, summing all such components gives

$$\bar{H}(j\omega) = j\omega\tau \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \right)^2 \left[\sum_{k=-\infty}^{\infty} H(k\tau) e^{-jk\omega\tau} \right] \quad \dots \quad (10)$$

Comparing this expression with eqn. (5) one sees following the argument of eqns. (6) and (7) that the relative error due to the finiteness of τ is for the straight-line approximation

$$|\Delta\tau| \simeq 1 - \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \right)^2 \quad \dots \quad (11)$$

(3.3) Mid-Interval Straight-Line Approximation

Since the errors introduced by the straight-line approximation of the previous Section, given by eqn. (11), are about twice as large as those for mid-interval stepped approximations, it seems worth while to examine another straight-line approximation in which the centres of the straight-line segments rather than the ends coincide with the system indicial response (see Fig. 7).

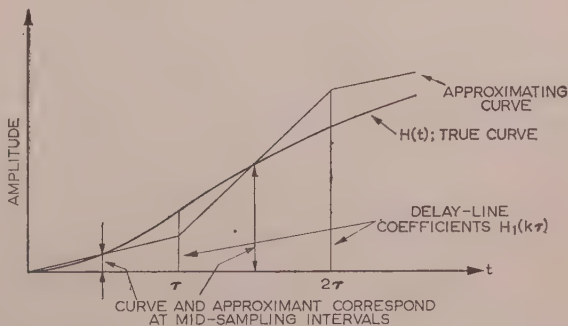


Fig. 7.—Output waveform for mid-interval straight-line approximation.

Similar to eqn. (10), the transfer function of this approximation for an infinite delay-line synthesizer is

$$H(j\omega) = j\omega\tau \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau}\right)^2 \left[\sum_{k=-\infty}^{\infty} H_1(k\tau) e^{-jk\omega\tau} \right] \quad (12)$$

But now the centres of straight-line segments are required to coincide with the system indicial response, i.e.

$$\frac{1}{2} [H_1(k\tau) + H_1(k\tau + \tau)] = H(k\tau + \frac{1}{2}\tau) \quad (13)$$

from which follows

$$H_1(k\tau) = 2 \sum_{l=0}^{\infty} (-1)^l H(k\tau - l\tau) \quad (14)$$

Introducing eqn. (14) into eqn. (12) and making use of

$$e^{-\frac{1}{2}j\omega\tau} \sum_{l=0}^{\infty} (-1)^l e^{-j\omega l\tau} = \frac{1}{2 \cos \frac{1}{2}\omega\tau}$$

we have

$$\bar{H}(j\omega) = j\omega \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \frac{\tan \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \sum_{l=0}^{\infty} H(k\tau + \frac{1}{2}\tau) e^{-j(k+\frac{1}{2})\omega\tau}$$

from which it can be concluded that for this approximation the relative error is given by

$$|\Delta\tau| \simeq \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \frac{\tan \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} - 1 \quad (15)$$

For small values of $\omega\tau$ the error of this approximation is almost the same as for the mid-interval stepped approximation (see Table 1).

(3.4) Curved Approximation

Table 1 and Fig. 8 illustrate the magnitudes of errors for various values of $\omega\tau$ introduced by the weighting functions of the types of approximation so far considered, together with those

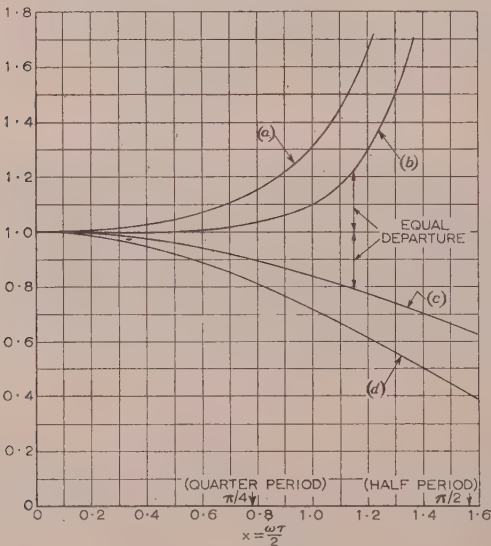


Fig. 8.—Sampling weighting functions.

- (a) $\frac{\sin x}{x} \frac{\tan x}{x}$ straight-line mid-interval approximation.
- (b) $\left(\frac{\sin x}{x}\right)^2 \frac{\tan x}{x}$ curved-line mid-interval approximation.
- (c) $\frac{\sin x}{x}$ mid-interval stepped approximation.
- (d) $\left(\frac{\sin x}{x}\right)^2$ straight-line approximation.

Table 1
PERCENTAGE DEPARTURES FROM UNITY

Type of approximation	Weighting function*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.4	1.57
(a) Eqn. (8)	$\frac{\sin x}{x}$	-0.17	-0.67	-1.50	-2.65	-4.11	-5.89	-7.97	-10.33	-12.96	-15.85	-22.33	-30	-36.31
(b) Eqn. (11)	$\left(\frac{\sin x}{x}\right)^2$	-0.34	-1.34	-2.96	-5.23	-8.05	-11.43	-15.31	-19.59	-24.24	-29.19	-39.67	-50.4	-59.44
(c) Eqn. (15)	$\left(\frac{\sin x}{x}\right)\left(\frac{\tan x}{x}\right)$	+0.16	+0.68	+1.57	+2.90	+4.77	+7.30	+10.74	+15.40	+21.87	+31.5	+66.45	+190	
(d) Eqn. (16)	$\left(\frac{\sin x}{x}\right)^2\left(\frac{\tan x}{x}\right)$	+0.00	+0.00	+0.06	+0.17	+0.46	+0.99	+1.91	+3.49	+6.08	+10.28	+29.29	+105.1	

(a) Stepped mid-interval approximation.
(b) Straight-line approximation.
(c) Straight-line mid-interval approximation.
(d) Curved-line approximation.
* $x = \omega\tau/2$.

of a further one which is markedly superior and which will now be considered. Comparison of the first and third rows of Table 1 shows that for small values of $\omega\tau$ the percentage errors are very similar. This is a consequence of $(\tan \frac{1}{2}\omega\tau)/\frac{1}{2}\omega\tau$ exceeding unity by almost exactly twice as much as $(\sin \frac{1}{2}\omega\tau)/\frac{1}{2}\omega\tau$ is less than unity. Thus, since the departures from unity are small the errors would be much smaller if $\Delta\tau$ were represented by

$$\Delta\tau \simeq \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau}\right)^2 \frac{\tan \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} - 1 \quad (16)$$

The final row in Table 1 gives the errors for such a case which lie within 4% up to $\omega\tau/2 = \pi/4$, the quarter-period (see Fig. 8).

In order to be able to employ this weighting function one seeks to construct an approximation such that a formula for $\bar{H}(j\omega)$ is derived (except for a minor change) from eqn. (12) by multiplying by a factor $\sin \frac{1}{2}\omega\tau/\frac{1}{2}\omega\tau$, which can then be matched to the required response at mid-interval points using eqn. (13). In this way one is led to an approximation based upon the curve shown in Fig. 9 instead of the isosceles triangle in Fig. 6.

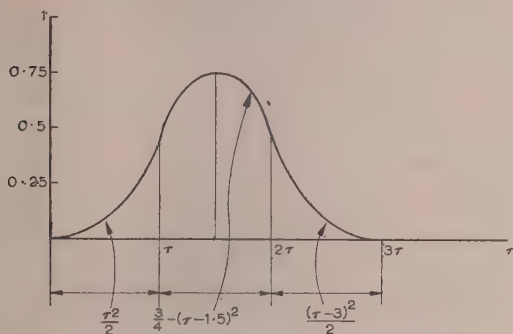


Fig. 9.—Basic indicial response for the curved approximation.

The transfer function of a system with this indicial response is given by

$$\bar{H}_c(j\omega) = j\omega\tau \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau}\right)^3 e^{-\frac{1}{2}j\omega\tau} \quad (17)$$

which is (except for a shift factor) equal to the transfer function (10) corresponding to the isosceles triangle multiplied by $\sin \frac{1}{2}\omega\tau/\frac{1}{2}\omega\tau$. An infinite delay-line synthesizer in which the summed potential in Fig. 1 has been operated on by a device having the above indicial response, and in which the first sample has no delay, has a transfer function given by

$$\bar{H}(j\omega) = j\omega \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau}\right)^3 \left[\sum_{k=-\infty}^{\infty} a_k e^{-j(k+\frac{1}{2})\omega\tau} \right] \quad (18)$$

The form of the basic indicial response shown in Fig. 9 is fortunately such that the indicial response of this delay-line synthesizer at any sampling moment $k\tau$ is $\frac{1}{2}(a_{k-1} + a_k)$, and one can therefore ensure that

$$\frac{1}{2}(a_{k-1} + a_k) = H(k\tau) \quad (19)$$

From eqns. (18) and (19) it follows that for this delay-line synthesizer the required weighting function of eqn. (16) indeed obtains in the same manner as eqn. (15) follows from eqns. (12) and (13).

However, it does not seem practical to use a device with the indicial response shown in Fig. 9 in an actual delay line, which must be finite since its indicial response would become zero after a finite time. Fortunately, the basic response shown in Fig. 9 has discontinuities in its second derivative at τ and 2τ , while

elsewhere it is constant. Consequently, the imitated response will have discontinuities in second derivative at sampling points only. Hence the practical realization of this form of approximation is by means of a tapped delay line followed by two integrators. The approximation must be made to coincide at the points of discontinuity in the second derivative (being the sampling moments) with the system indicial response, as is illustrated in Fig. 10.

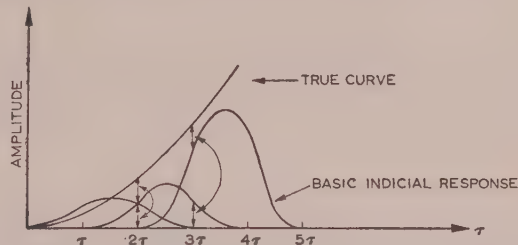


Fig. 10.—Analysis of output waveform for the curved approximation. Approximation is given by the sum of basic responses.

It is perhaps worth commenting that the basic indicial response used for the purposes of analysing this last form of approximation is a most favourable one for use in practical pulse sampling in general and could with advantage be exploited for this purpose.

(4) THE ORDER OF ERRORS

In order to give some idea of the scale of errors involved in the analogue representation, a preliminary calculation is made of the average error over a working band of frequencies for a representation of a simple 3-time-constant system. The working band of frequencies is taken to be a quarter of the periodic interval arising from the sampling time τ . The three time-constants are assumed to be equal, so that the system transfer function is given by

$$H(j\omega) = \frac{1}{\left(1 + \frac{j\omega}{3}\right)^3} \quad (20)$$

The impulse response of the system is

$$h(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{\left(1 + \frac{p}{3}\right)^3} e^{pt} dp = \frac{3^3 t^2 e^{-3t}}{2} U(t) \quad (21)$$

whence the indicial response of the system is given by

$$H(t) = \int_{-\infty}^t h(t) dt \quad (22)$$

The 3-time-constant system is the simplest that can become unstable with direct proportional control, and the frequency at which it first becomes unstable is that for which it has 180° phase lag in response to sinusoidal input. The proportional gain for sustained oscillation will be greater with the 3-time-constant system than with systems involving a greater number of equal time-constants, for the reason that the modulus frequency characteristic falls off with increasing phase more rapidly than any other case. The 3-time-constant system may thus be regarded as inherently more stable than higher-order systems, and as such involves a broader working band of frequencies.

However, with a 3-time-constant system an even broader band of frequencies will be involved if the system is associated with a phase-advance network. In this case the frequency at which the

system goes into self-oscillation will be that for which the system alone has a phase lag exceeding 180° by as much as, say, 60° , i.e. a phase lag of 240° . This, it may be assumed, is the worst case that will arise from the point of view of a broad band of frequencies being involved, and so, for the present calculation, the working band of frequencies will be taken as from zero up to the frequency for which the system has 240° phase lag. This frequency will be the one for which each of the three time-constants contributes 80° phase lag, i.e.

$$\frac{\omega_1}{3} = \tan 80^\circ \text{ or } \omega_1 = 17 \text{ rad/sec}$$

But this frequency is to be the quarter-periodic frequency arising from the sampling time τ , i.e. $\omega_1 = \pi/2\tau$, whence $\tau = 0.0925$ sec.

(4.1) Errors due to the Delay-Section Time, τ

With the system transfer function given, the assumption that overlapping contributions from the other periods of the spectrum are negligible ($<1\%$ in the mean) is justified, provided that the working band of frequencies is confined to the quarter-period range $0 < \omega < \pi/2\tau$, and hence expression (8) may be used in the case of the mid-interval stepped approximation to assess error due to the delay-section time τ .

Hence the mean error over the working band of frequencies is given by

$$\begin{aligned} \frac{1}{\omega_1} \int_0^{\omega_1=17} \left(1 - \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau}\right) d\omega &= \frac{4}{\pi} \int_0^{\pi/4} \left(1 - \frac{\sin x}{x}\right) dx \\ &= \frac{4}{\pi} \left(\frac{\pi}{4} - \int_0^{\pi/4} \frac{\sin x}{x} dx\right) = 0.0336 \text{ or } 3.36\% \end{aligned}$$

This compares favourably with the case where the whole of the available half-period range of frequencies is employed as the working band corresponding to $\omega_1 = 17 = \pi/\tau$, or $\tau = 0.185$ sec.

So far as the accuracy of representation of the high frequencies is concerned the half-period case is much worse. Thus for the quarter-period case the percentage error at ω_1 is

$$\left(1 - \frac{\sin \frac{\pi}{4}}{\frac{\pi}{4}}\right) 100 = 10\%$$

the overlap from higher harmonics being negligible; while for the half-period case it is 36% plus a further 36% for overlap from the next harmonic. It is quite clear that it is not possible to make full use of the half-period range available if a good representation is to be obtained at the higher frequencies.

(4.2) Error due to the Finite Number of Delay Sections, N

This is the error that arises as a result of the system indicial response being no longer imitated in detail in the approximation after some time $T = N\tau$. The true system frequency-transfer function is given by

$$H(j\omega) = \int_0^\infty h(t) e^{-j\omega t} dt \quad (23)$$

where $h(t)$ is the system impulse response. When $h(t)$ is terminated at time T , the approximating frequency function $\bar{H}(j\omega)$ is given by

$$\bar{H}(j\omega) = \int_0^T h(t) e^{-j\omega t} dt \quad (24)$$

The error, being the difference between eqns. (23) and (24), is given by

$$\int_T^\infty h(t) e^{-j\omega t} dt \quad (25)$$

Substitution for the impulse response $h(t)$ of the system given in eqn. (21) and integration gives

$$\text{Error} = \left[\frac{\frac{9}{2}T^2}{\left(1 + \frac{j\omega}{3}\right)} + \frac{3T}{\left(1 + \frac{j\omega}{3}\right)^2} + \frac{1}{\left(1 + \frac{j\omega}{3}\right)^3} \right] e^{-(3+j\omega)T} \quad (26)$$

$$\text{or} \quad |\text{Error}| \approx \frac{9}{2} \frac{T^2 e^{-3T}}{\left(1 + \frac{\omega^2}{9}\right)^{1/2}}$$

$$\text{when} \quad T\left(1 + \frac{j\omega}{3}\right) \gg 1$$

The true frequency function has modulus

$$|H(j\omega)| = \frac{1}{\left(1 + \frac{\omega^2}{9}\right)^{3/2}}$$

Thus the modulus of error as a proportion of $|H(j\omega)|$ is given by

$$\frac{|\text{Error}|}{|H(j\omega)|} = \frac{9}{2} T^2 e^{-3T} \left(1 + \frac{\omega^2}{9}\right) \quad (27)$$

The mean value of this ratio over the frequency range of interest is given by

$$\frac{9}{2} \frac{T^2}{\omega_1} e^{-3T} \int_0^{\omega_1=17} \left(1 + \frac{\omega^2}{9}\right) d\omega = \frac{9}{2} \frac{T^2}{17} e^{-3T} \left[\omega + \frac{\omega^3}{27}\right]_0^{17} = 52.6 T^2 e^{-3T}$$

for $T = 2.9$ sec, mean error = 0.072 or 7.2% .

The magnitude of error at $\omega = 17$, by substitution in eqn. (26), is 1.10×10^{-3} in an ordinate $|H(j\omega)|_{\omega=17}$ of 5.22×10^{-3} ; that is a relative error of about 20% . This is already rather large, but as is shown by relation (27), the relative error increases as the square of frequency for $\omega \gg 3$. Thus in this case the value of T is the dominant factor and the magnitude of error depends fairly critically on its value.

Taking both sources of error together on the assumption that the errors are directly additive gives an overall accuracy in terms of mean values of about 10% for the delay interval $\tau = 0.0925$ sec and the total delay time $T = 2.9$ sec. The number of sections in the delay line, $N = T/\tau$, is 31 for this case, which represents the most difficult one that can occur from the aspect of frequency bandwidth. It remains to be seen what proportion of practical cases will be so demanding as this, but it seems probable that a very much cruder delay-line will serve for the majority of cases. The figure given for the number of sections represents the upper limit of what would normally be necessary and serves to indicate that the method of representation is not impractical. In the next Section results are given of some detailed computations on delay-lines having various numbers of sections up to 31 used to

represent two system transfer functions chosen as being representative of the normal type of system met with in practice.

(5) DETAILED COMPUTATIONS FOR TWO SYSTEMS

For the purposes of making detailed numerical calculations, systems were considered having a greater number of exponential delays than the previous case of three. Systems were chosen having four and six equal time-constants, as being more representative of practical systems, i.e.

$$H(j\omega) = \frac{1}{\left(1 + \frac{j\omega}{4}\right)^4} \text{ and } H(j\omega) = \frac{1}{\left(1 + \frac{j\omega}{6}\right)^6}$$

The notation used is as follows:

- Number of time-constants, n , = 4 or 6.
- Sampling interval (time), τ , = 0.1 sec.
- Number of delay-line sections, N , = 20, 26, 31.
- Frequency interval, Δ , = $\pi/2$ rad/sec.
- Number of frequency intervals, h , = 1-6.

The transfer functions, $\bar{H}(j\omega)$, approximating to these systems were calculated using the formula

$$\bar{H}(j\omega) = 2j \sin \frac{1}{2}\omega\tau \sum_{k=0}^{N-1} H(k\tau)e^{-j\omega k\tau} + e^{-j\omega(N-\frac{1}{2})\tau} \quad (28)$$

which is analogous to eqn. (4) but applies to the mid-interval stepped approximation. These functions were evaluated for a delay-line of 20 sections ($N=20$) having sampling intervals $\tau = 0.1$ sec at frequency intervals $\Delta = \pi/2$ rad/sec, there being h such intervals. It was thus possible to determine the magnitude of error at each of the h frequencies; this error is strictly a vector quantity, being the difference between two complex numbers, but in the first instance attention is given to its magnitude only.

For a delay line consisting of more than 20 sections it was found more convenient to build on the results for a 20-section line up to the required number of sections by means of a formula expressing the change due to the additional sections only. This formula is derived as follows:

For a line of N_1 sections error arises due to the truncated form of the response, since $H(t)$ is replaced by unity for $t > N_1\tau$:

$$\text{Error} = 2j \sin \frac{1}{2}\omega\tau \sum_{k=N}^{\infty} e^{-jk\omega\tau} - 2j \sin \frac{1}{2}\omega\tau \sum_{k=N}^{\infty} H(k\tau)e^{-jk\omega\tau}$$

i.e. the error is accounted for by the difference between $H(t)$ and unity for $t > N\tau$. When the delay line is extended only to, say, N_2 sections, the differences in error between the two cases are given by

$$\text{Modification to } \bar{H}(j\omega) \text{ for } (N_2 - N_1) \text{ sections} \\ = 2j \sin \frac{1}{2}\omega\tau \sum_{N_1}^{N_2-1} [1 - H(k\tau)]e^{-j\omega k\tau} \quad (29)$$

The frequencies of greatest interest are those for which the complete system is likely to become unstable. If these are represented accurately, lower frequencies in the normal working band will be more accurately represented. Much here depends on the form of the stabilizing network, but two cases have been considered: one in which the network provides very little phase advance and the other in which it provides a maximum phase advance of 50° . The frequencies at which self-oscillation would occur have been calculated for both of the systems for each of these cases and are given below.

(a) Fourth-Order System ($n=4$).

Basic system transfer function has 198° phase lag at $\omega = 3\pi/2$ ($h=3$) (i.e. 18° phase advance).

Basic system transfer function has 230° phase lag at $\omega = 2\pi$ ($h=4$) (i.e. 50° phase advance).

(b) Sixth-Order System ($n=6$).

Basic system transfer function has 199° phase lag at $\omega = 5\pi/4$ ($h=2\frac{1}{2}$).

Basic system transfer function has 229° phase lag at $\omega = 3\pi/2$ ($h=3$).

Table 2

ERRORS IN THE TRANSFER FUNCTION EXPRESSED IN PERCENTAGE MODULUS AT 198° PHASE LAG OF THE BASIC SYSTEM

	Fourth-order system			Sixth-order system		
	$N=20$	$N=26$	$N=31$	$N=20$	$N=26$	$N=31$
Error due to truncation	%	%	%	%	%	%
Error due to truncation	10	3	0.63	3.7	0.14	0.0
Error due to stepping*	0.93	0.93	0.93	0.64	0.64	0.64
Vector sum for mid-interval stepped approximation..	11	3.7	1.2	4.1	0.51	0.64
Vector sum for straight-line approximation	12	4.35	1.95	4.6	1.15	1.28
Percentage magnitude of last step a_k ; $k=N$	4.3	0.7	0.16	2.0	0.18	0.021

* Double for straight-line approximation.

Table 3

ERRORS IN TRANSFER FUNCTION EXPRESSED IN PERCENTAGE MODULUS AT 230° PHASE LAG OF THE BASIC SYSTEM

	Fourth-order system			Sixth-order system		
	$N=20$	$N=26$	$N=31$	$N=20$	$N=26$	$N=31$
Error due to truncation	%	%	%	%	%	%
Error due to truncation	52	7.1	1.9	5.54	0.51	0.04
Error due to stepping	1.64	1.64	1.64	0.93	0.93	0.93
Vector sum for mid-interval stepped approximation..	53.5	5.7	3.5	5.36	1.0	0.95
Vector sum for straight-line approximation	55	4.5	5.3	5.4	1.9	1.86

The highest of these frequencies has $h = 4$, and from Table 1 it will be observed that the value chosen for τ , namely 0.1 sec, gives a stepping error due to a mid-interval approximation of 1.6% for $h = 4$ with an infinite delay-line. Using formulae (28) and (29) the total error for a delay line of a finite number of sections may be calculated; the results are summarized in Tables 2 and 3 for delay lines having 20, 26, and 31 sections.

From Tables 2 and 3 it will be observed that a very considerable improvement in accuracy occurs for a delay line of a given number of sections as between a fourth- and a sixth-order system. This improvement can be attributed to two factors, namely an increase in setting discrimination in view of the small size of coefficients a_k ($k \rightarrow N$) implied, and a greater relative modulus of transfer function at a given frequency.

The following figures show the order of improvement that might be expected if this is the case:

$$\left. \begin{array}{l} \text{For } n = 4 : [1 - H(t)]_{t=3.1} = 0.00163 = a_k; \quad k = 31 \\ \text{For } n = 6 : [1 - H(t)]_{t=3.1} = 0.000207 \end{array} \right\} \text{ratio } 7.8$$

$$\left. \begin{array}{l} \text{For } n = 4 : |H(j\omega)|_{190^\circ \text{ phase}} = 0.175 \\ \text{For } n = 6 : |H(j\omega)|_{190^\circ \text{ phase}} = 0.343 \end{array} \right\} \text{ratio } 2$$

These effects are cumulative, so that the net ratio is 15.6.

The manner in which the approximating frequency-response vectors approach their true values as the number of delay-line sections increases is interesting. The process is illustrated in Fig. 11, which shows the time vectors for a fourth-order process for $h = 2, 3$, and 4, and the spiral approach of the approximating vectors as the number of sections increases. It will be noted that the spirals have more revolutions as the frequency increases (h increasing) and that the spirals terminate at points not corresponding with the true vectors. The outstanding difference represents the error due to stepping with an infinite line. The three vectors and their approximants are shown more precisely and in greater detail in Fig. 12, from which it will be seen that the incremental improvement per section decreases as the number of sections is increased.

The accuracies so far quoted have referred to frequency-response vectors at particular frequencies and express the degree to which the approximating vectors are in error. In many aspects of system analysis for which an analogue of the type being discussed is useful, a more important consideration is to know the change in frequency that must be made before the true system transfer function has the same phase lag as one of its approximants. In other words, how serious are the errors in terms of frequency departures instead of magnitude and phase departures as has so far been assessed?

An estimate of errors on a frequency basis can be made for a particular vector approximant by taking the phase error and dividing by the rate of change of phase with frequency:

$$\text{Error in frequency} = + \frac{\text{error in phase}}{\partial \phi / \partial \omega}$$

where ϕ is the phase angle.

Table 4 gives a comparison between errors assessed on a

Table 4

COMPARISON OF ERRORS FOR A FOURTH-ORDER SYSTEM ($n = 4$).

Number of sections, N	Frequency number, h	Frequency error	Modulus error
26	2	%	%
26	3	0.76	1.3
		4.2	3.7
31	2	0.0	0.1
31	3	0.75	1.2

modulus basis and errors assessed on a frequency basis in this fashion.

These figures show that errors in the frequency at which, for example, self-oscillation occurs in a system based on a delay-line

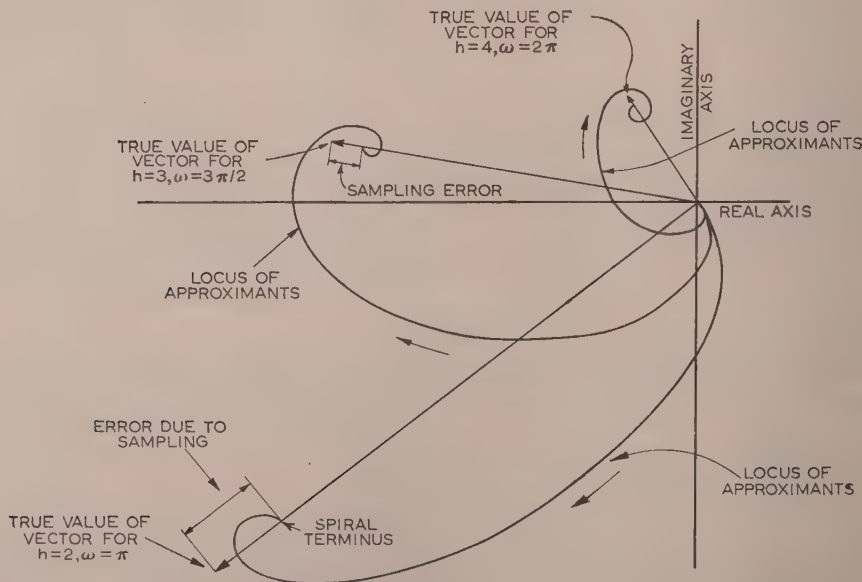


Fig. 11.—Influence of numbers of delay-line sections on the approximating frequency response vectors.

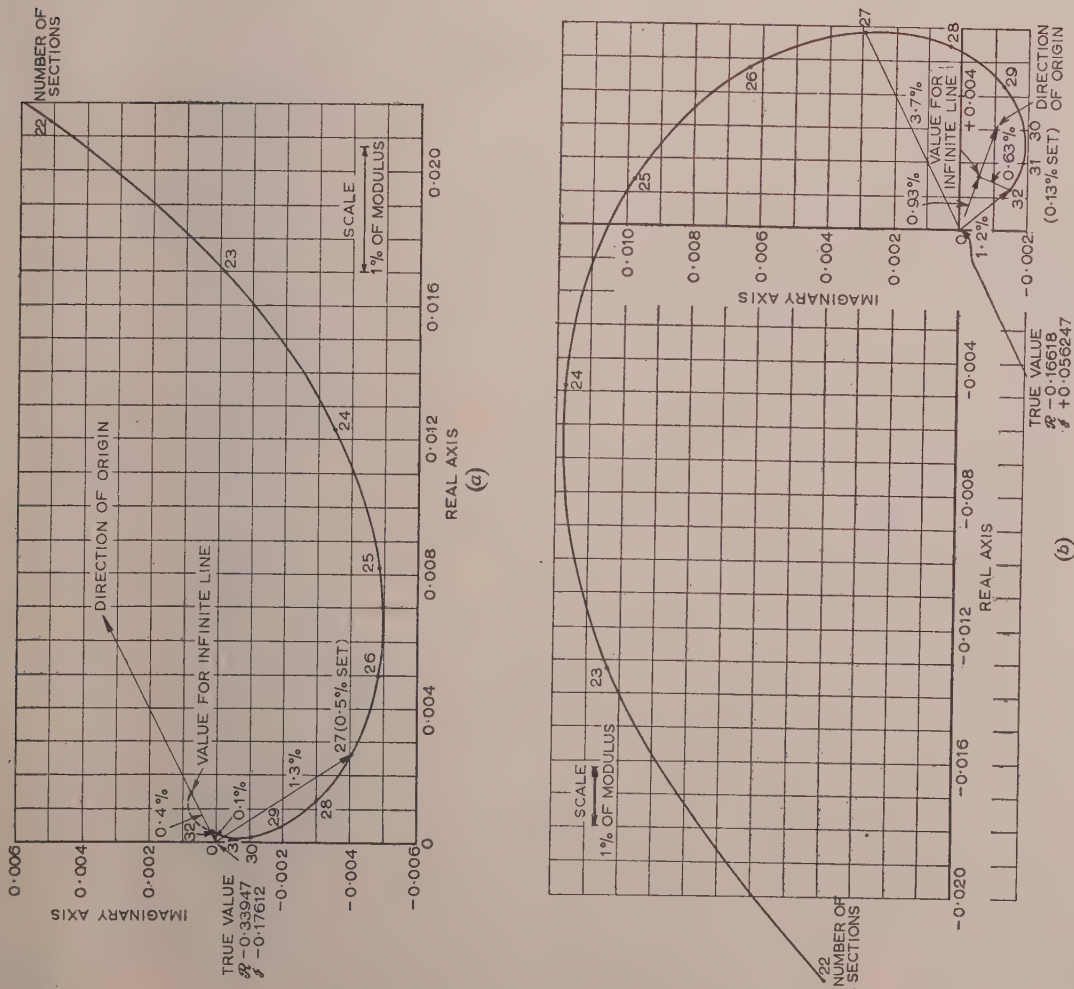


Fig. 12.—Transfer functions for varying number of sections.

analogue will be less than the corresponding modulus errors in the majority of cases. In view of this it appears that the modulus accuracy figures are fairly representative of the characteristics of interest.

(6) ACKNOWLEDGMENT

The author wishes to thank members of the Measurement and Control Department of the Royal Dutch-Shell Laboratory, Delft, for interesting him in this problem. Special mention should be made of the assistance received from R. M. F. Houtappel in numerous discussions during the preparation of the paper.

(7) REFERENCES

- (1) KALLMAN, H.: 'Transversal Filters', *Proceedings of the Institute of Radio Engineers*, 1940, 28, p. 302.
- (2) HEBB, M. H., HORTON, C. W., and JONES, F. B.: 'On the Design of Networks for Constant Time Delay', *Journal of Applied Physics*, 1949, 20, p. 616.
- (3) LAWREN, D. F.: 'A General Theory of Sampling Servo Systems', *Proceedings I.E.E.*, Monograph No. 4, July, 1951 (98, Part IV, p. 31).
- (4) MAGNUS, W., and OBERHETTINGER, H.: 'Formeln und Satze' (Springer, Berlin, 1948), second edition, p. 217.

PROPERTIES OF A FEEDBACK-SYSTEM ANALOGUE BASED ON A DISCONTINUOUS DELAY-LINE SYNTHESIZER

By R. M. F. HOUTAPPEL, Dr.Ir.

(The paper was first received 16th September, 1955, and in revised form 26th January, 1956. It was published as an INSTITUTION MONOGRAPH May, 1956.)

SUMMARY

The paper considers the frequency response of a feedback-system analogue in which the system is imitated by a discontinuous delay-line synthesizer. The various sources of error in the imitation transfer functions of the delay-line synthesizer and stabilizing network, as compared with the actual transfer functions of system and stabilizing network, are discussed. The contribution of each of these sources is calculated for a feedback system consisting of a multi-capacity process and a 3-term controller.

The results of these calculations are used to determine how to minimize each of these contributions. Graphs are included from which the number of delay-line sections required for a prescribed accuracy and the best adaptation of the time scale of the process to that of the delay line may be determined.

LIST OF PRINCIPAL SYMBOLS

- V = Voltage (for suffixes see Figs. 1 and 2).
- \bar{V} = Complex amplitude of voltage.
- α_k = Dimensionless constant (see Fig. 2).
- τ = Sampling period.
- σ = Change-over time interval.
- t_0 = Moment of step-function input (see Figs. 4 and 5).
- T = Time interval in which the indicial response of the system can be imitated in detail [see eqn. (20)].
- T_I = Time-constant of approximate integrator.
- T_i = Time-constant of integral action of stabilizing network.
- T_d = Time-constant of derivative action of stabilizing network.
- $H(t)$ = Indicial response of system.
- $U(t)$ = Unit step function ($=0$ for $t < 0$; $=1$ for $t > 0$).
- $H_A(t), H_B(t), H_C(t)$ = Indicial responses equal to $H(t)$ for $t < T$ but deviating from $H(t)$ for $t > T$ according to Fig. 4.
- Φ_p, Φ_{st} = Transfer function of system and stabilizing network respectively.
- $\Phi_{p,T}$ = Transfer function corresponding to indicial response $H_A(t), H_B(t)$, or $H_C(t)$.
- $\Phi_D^*, \Phi_{st}^*, \Phi_{circuit}^*$ = Imitation transfer function of delay-line synthesizer, stabilizing network and open-loop circuit, respectively (Φ_D^* and $\Phi_{circuit}^*$ are, in good approximation, clamping transfer functions).
- $\Phi_{\infty}^*, \Phi_{\infty,\sigma}^*$ = Imitation transfer function of infinite delay-line synthesizer considered as a pure clamping system with response $V_0(t)$ to input $V_1 = U(t - t_0)$ for which respectively $V_0(k\tau) = H(k\tau - t_0)$ and $V_0(\frac{1}{2}k\tau + \sigma) = H(\frac{1}{2}k\tau + \sigma - t_0)$.

- $2N$ = Total number of buffer amplifiers in delay line.
- P_4, P_5, P'_4 = System or process with transfer function given by eqns. (24), (25) and (26) respectively.
- Δ_T = Relative error due to the finite duration, T , of the detailed imitation.
- Δ_τ = Relative error due to the sampling interval, τ , with $V_0(k\tau) = H(k\tau - t_0)$.
- $\Delta_{\tau,\sigma/\tau}$ = Relative error due to sampling interval, τ , and switching interval, σ , with $V_0(\frac{1}{2}k\tau + \sigma) = H(\frac{1}{2}k\tau + \sigma - t_0)$.
- Δ_i, Δ_d = Relative errors in the imitation transfer function of the stabilizing network related to the integral and derivative actions respectively [see eqn. (43)].
- $\bar{\Delta}_{tot} = |\Delta_T| + |\Delta_{\tau,\sigma/\tau}|$.
- Δ' = Relative error in imitation transfer function of delay-line synthesizer due to restriction $\alpha_1 = \alpha_2 = 0$ (see Fig. 2).

(1) INTRODUCTION

As has been shown by Kallman,¹ a delay-line synthesizer consisting of a number of equal delaying units in cascade and a summing device can be used to imitate the indicial response (response to a step-function input) of a system. If the delaying units are formed by identical LC networks² the delay-line synthesizer is itself a continuous system. Its properties as a system analogue have been studied by Westcott.³ There are, however, technical reasons⁴ for preferring in this application a discontinuous delay-line consisting of buffer amplifiers, storage capacitors and switches.⁵

With a discontinuous delay-line synthesizer the indicial response of a system can be imitated in the same way as with a continuous one.³ In two respects, however, the behaviour of the discontinuous delay-line synthesizer so adjusted is more complicated than that of the continuous one. First, the choice of the moment at which the step-function input is introduced and the indicial response imitation starts is relevant in the discontinuous case but not in the continuous case. Secondly, in contrast with the case of the continuous delay-line, a sinusoidal input causes a non-sinusoidal output of the discontinuous delay-line synthesizer, which, if the delay-line synthesizer forms part of a feedback-system analogue (see Figs. 1A and 1B), affects the behaviour of the stabilizing network.

In the paper a simple method is presented of deciding how many sections the discontinuous delay-line synthesizer must have

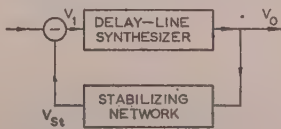


Fig. 1A.—Feedback-system analogue.

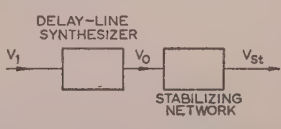


Fig. 1B.—Open loop of feedback-system analogue.

Correspondence on Monographs is invited for consideration with a view to publication.
Dr. Houtappel is at the Koninklijke/Shell-Laboratorium, Delft.

and how it may be used (by imitating the indicial response of the system) to approximate with a given accuracy the frequency characteristics of the imitated system considered as part of a feedback system. Particular attention is given to the imitation of a higher-order process (i.e. one having a large number of exponential time-constants) in a control loop with a controller fitted with proportional, integral and derivative actions. Interest centres on higher-order processes, since simpler processes are more easily represented directly by RC networks.

(2) PRINCIPLE OF THE DISCONTINUOUS DELAY-LINE SYNTHESIZER

(2.1) Description

Figs. 2A and 2B show the discontinuous delay-line synthesizer. Via the buffer amplifiers B_1 – B_{2N} , with amplifications equal to 1,

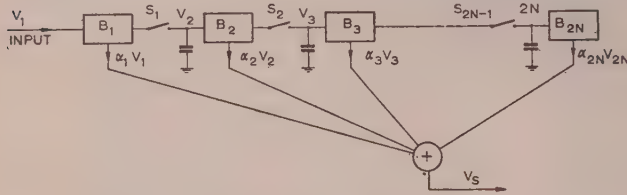


Fig. 2A.—Discontinuous delay-line synthesizer.

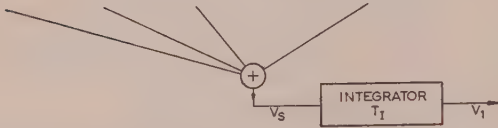


Fig. 2B.—Discontinuous delay-line synthesizer with integrator.

and the electronic switches S_1 – S_{2N-1} , the input potential $V_1(t)$ is passed on to the right. To this end all the odd-numbered and all the even-numbered switches are alternately made conductive. During the brief change-over time intervals neither the odd nor the even switches are conductive.

Fig. 3 shows how, for an arbitrary input voltage $V_1(t)$, the voltages V_2 , V_3 , etc., change with time. Because, as can be seen in Fig. 3, just those values of the input voltage $V_1(t)$ are passed

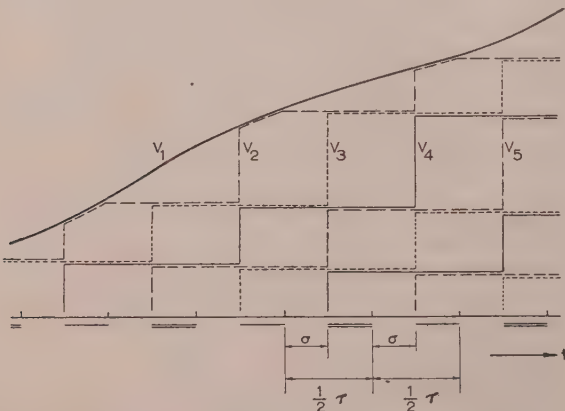


Fig. 3.—The potentials $V_k(t)$ for an arbitrary $V_1(t)$.

The intervals during which the odd and even switches are conductive are indicated beneath the time axis by a single and a double bar respectively.

τ = Sampling period.
 σ = Change-over time interval.

on that occur at the moments that the odd-numbered switches become non-conductive, we shall call these moments the sampling moments. The time interval between two consecutive sampling moments is constant and will be called the sampling period, τ . The time interval between a sampling moment and the subsequent moment when the even-numbered switches become non-conductive is also constant and is equal to $\frac{1}{2}\tau$.

It is, in general, technically difficult to keep the change-over time intervals, σ , constant. However, to calculate their influence to a first approximation we assume them to be so ($0 < \sigma < \frac{1}{2}\tau$). This assumption simplifies the calculations considerably.

From now on we choose as the time zero one of the sampling moments. We see that, if k and s are whole numbers and $s \geq 0$,

$$V_2(t) = V_1(k\tau) \text{ for } k\tau < t < (k + \frac{1}{2})\tau + \sigma \quad (1)$$

$$V_2(t) = V_1(t) \text{ for } (k + \frac{1}{2})\tau + \sigma < t < (k + 1)\tau \quad (2)$$

$$V_{2s+3}(t) = V_1(k\tau - s\tau) \text{ for } k\tau + \sigma < t < (k + 1)\tau + \sigma \quad (3)$$

$$V_{2s+4}(t) = V_1(k\tau - s\tau) \text{ for } (k + \frac{1}{2})\tau + \sigma < t < (k + 1\frac{1}{2})\tau + \sigma \quad (4)$$

From the voltages $V_r(t)$, fractions $\alpha_r V_r(t)$ are formed by means of potentiometers (α_r can be positive or negative). These fractions are added in a summing device and yield the sum voltage

$$V_S = \alpha_1 V_1 + \alpha_2 V_2 + \dots + \alpha_{2N} V_{2N} \quad (5)$$

Throughout the paper the number of buffer amplifiers is assumed to be even. Calculations for an odd number yield essentially the same results. Because between two consecutive sampling moments the value of V_k is passed on to V_{k+2} , N and not $2N$ is to be compared with the number of sections of a continuous delay line.

(2.2) Methods of Imitating a System

The easiest method of imitating a system by a discontinuous delay-line synthesizer is to make the indicial response of the delay-line synthesizer visible on an oscillograph and, by adjusting the potentiometers, to make it similar to the indicial response of the system. To this end one equates the input of the delay line to

$$V_1(t) = U(t - t_0) \quad (6)$$

with $U(x)$ the unit step function and t_0 a small negative time, $-\tau < t_0 < 0$, for which the best choice—as discussed in Section 3.2—appears to be $-\frac{1}{2}\tau$. The output, V_0 , is now made similar to $H(t - t_0)$.

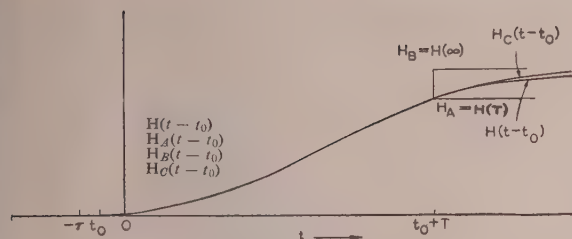
Because of the special importance of the sampling moments ($t = k\tau$), one must attempt as far as possible to equate the two indicial responses for these moments

$$V_0(k\tau) = H(k\tau - t_0) \quad (7)$$

It follows from eqns. (1)–(5) that for an input given by eqn. (6) the choice of the output of the delay-line synthesizer, if $V_0 = V_S$, is subject to the restriction

$$V_0(t) = \text{constant} = V(\infty) \text{ for } t > (N - \frac{1}{2})\tau + \sigma \quad (8)$$

whereas the values of $V_S(k\tau)$ for $0 \leq k \leq N - 1$ can be chosen arbitrarily and independently of each other. Restriction (8) causes a rather serious imperfection in the resulting frequency response of the system; this imperfection is, of course, dependent on the choice of the constant $V(\infty)$ in eqn. (8). Two different choices reduce to the imitation, respectively, of $H_A(t - t_0)$ and $H_B(t - t_0)$ shown in Fig. 4, instead of that of $H(t - t_0)$ by $V_0(t)$ according to eqn. (7).

Fig. 4.—Imitations of $H(t - t_0)$ for $t - t_0 > T$.

It is, however, not possible to study the influence of a derivative action of the stabilizing network in a feedback system with an electrical analogue in which the system is imitated by a discontinuous delay-line synthesizer in the above way. The output of the derivative action would, in general, reach saturation at the moments $t = \frac{1}{2}k\tau + \sigma$, because of the discontinuities in $V_S(t)$ at those moments, and would be zero—except for the small contribution of V_1 and V_2 —for all other times, and thus also for the sampling moments $t = k\tau$. The derivative action of the stabilizing network would therefore make almost no contribution to its imitation transfer function as defined in Section 3.1.

Therefore it seems reasonable to try to equate the properties of the discontinuous delay-line synthesizer more closely to those of the system by applying an integrator (see Fig. 2b) by which $V_S(t)$ is transformed into $V_I(t)$. The output, $V_0(t) = V_I(t)$, is then continuous for all times. By use of an approximate integrator for which

$$\frac{dV_I(t)}{dt} + \frac{1}{T_I}V_I(t) = V_S(t) \quad (9)$$

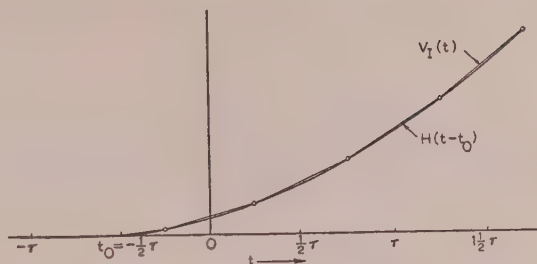
with adjustable time-constant, T_I , a second very important advantage is obtained. By adjusting T_I one can imitate the last part of the indicial response by a first-order indicial response having the same initial derivative and the same final value (imitation of H_C in Fig. 4); this, of course, results in a much better approximation than eqn. (8) (imitations of H_A and H_B in Fig. 4).

The rest of the paper will deal with the case in which the discontinuous delay-line synthesizer is provided with an integrator with adjustable time-constant, because of the obvious superiority of this application as a system analogue.

The settings of the delay-line synthesizer after the time-constant T_I has been chosen are not completely defined by the coincidence of $V(t)$ with $H(t - t_0)$ at the sampling moments $t = k\tau$ ($k \leq N - 1$) for a step input $V_1(t) = U(t - t_0)$. Indeed, this could even be achieved with only the odd-numbered or with only the even-numbered potentials $\alpha_k V_k$.

Now the sampling moments $t = k\tau$ are not indicated on the oscillograph, but the moments $t = k\tau + \sigma$ and $t = (k + \frac{1}{2})\tau + \sigma$ are recognizable from the discontinuities in the derivative of $V_0(t)$. It is therefore much more practicable, for an input (6), to aim at the fulfillment of $V_0(t) = H(t - t_0)$ for one or both of these last series of moments than for the sampling moments $t = k\tau$. The best approximation of $V(k\tau) = H(k\tau - t_0)$ will be achieved by doing this for both series of moments (see Fig. 5); thereby the settings of the delay-line synthesizer are defined completely for a prescribed value of T_I and all α 's must, in general, be used. After correct reduction of the time scale and choice of t_0 (see Section 4) the value of T_I is defined uniquely by the imitation of H_C (see Fig. 4).

Assuming $T_I > \tau$, which will, in general, be the case (see Section 4 and Fig. 15), the derivative of the output $V_0(t)$ is nearly constant between the two consecutive moments, $t = (k - \frac{1}{2})\tau + \sigma$

Fig. 5.—Imitation of $H(t - t_0)$ by $V_I(t)$ for $t - t_0 < T$.

$$V_I(\frac{1}{2}k\tau + \sigma) = H(\frac{1}{2}k\tau + \sigma - t_0) \text{ with } t_0 = -\frac{1}{2}\tau \text{ and } \sigma = \frac{1}{2}\tau.$$

and $t = k\tau + \sigma$, between which the sampling moment $t = k\tau$ lies, and closely approximates the derivative of $H(t - t_0)$ at $t = (k - \frac{1}{2})\tau + \sigma$. To make the derivative of $V_0(t)$ at the sampling moments $t = k\tau$ approximately equal to that of $H(t - t_0)$ at these same moments, the best choice of σ is therefore $\frac{1}{4}\tau$ (see Fig. 5).

(3) THEORY

(3.1) Imitation Transfer Functions of Discontinuous Delay-Line Synthesizer and Stabilizing Network

Fig. 3 and eqns. (1)–(4) show that the behaviour of $V_1(t)$ and $V_2(t)$ is markedly different from that of the other potentials $V_k(t)$ ($k \geq 3$). As to the contributions to V_S of V_3 , V_4 , etc., the discontinuous delay-line synthesizer is a pure clamping system,⁶ but the contributions of V_1 and V_2 spoil this property. A considerable simplification of the theory is achieved by neglecting this exceptional behaviour of V_1 and V_2 . This is justified by the fact that this effect is certainly smaller than that which a complete omission of V_1 and V_2 , i.e. the restriction

$$\alpha_1 = \alpha_2 = 0 \quad (10)$$

would have. Although it would then be impossible to imitate the very beginning of the indicial response of the system, even this, as will appear in Section 4 (Fig. 13), would make only a small contribution to the error in the imitation of the frequency response of a higher-order system.

From Fig. 3 or eqns. (1)–(5) can be deduced the following properties of the discontinuous delay-line synthesizer which obtain strictly only under restriction (10):

- The output depends on the input at the sampling moments $V_1(k\tau)$ only, and not on $V_1(t \neq k\tau)$.
- $V_S(t)$ remains constant during the time intervals $k\tau + \sigma < t < (k + \frac{1}{2})\tau + \sigma$ and $(k + \frac{1}{2})\tau + \sigma < t < (k + 1)\tau + \sigma$.

In the electric analogue of a feedback system in which the system is imitated by a discontinuous delay-line synthesizer the latter's output, $V_0(t)$, forms the input of the stabilizing network, and the output, $V_S(t)$, of this stabilizing network contributes to the input of the delay line (see Fig. 1A).

Because of property (a), only the values of the output of the stabilizing network at the sampling moments, $V_S(k\tau)$, are important. If the input of the delay line is sinusoidal, or—what is the only thing that matters (neglecting the exceptional behaviour of V_1 and V_2)—if at least at the sampling moments it can be described by

$$V_1(k\tau) = \Re(\bar{V}_1 e^{j\omega k\tau}) \quad (11)$$

with $\Re(x)$ the real part of the complex number x and \bar{V}_1 a complex constant, neither $V_0(t)$ nor $V_S(t)$ is, in general, purely sinusoidal. It can, however, easily be shown (see Section 6.1) that the values of $V_0(t)$ and $V_S(t)$ at the sampling moments,

$t = k\tau$, can then be exactly described by pure sinusoidal time functions having the same angular frequency, ω , as the input $V_1(t)$:

$$V_0(k\tau) = \mathcal{R}(\bar{V}_0 e^{j\omega k\tau}) \quad (12)$$

$$V_{St}(k\tau) = \mathcal{R}(\bar{V}_{St} e^{j\omega k\tau}) \quad (13)$$

The stabilizing properties of the circuit, not being influenced by the values of $V_{St}(t)$ at times $t \neq k\tau$, are now completely described by the ratio

$$\Phi_{circuit}^*(j\omega) = \bar{V}_{St}/\bar{V}_1 \quad (14)$$

which depends on the angular frequency ω and which will be called the imitation transfer function of the open-circuit; here it is virtually a clamping transfer function.[†] We can write

$$\Phi_{circuit}^*(j\omega) = \Phi_D^*(j\omega)\Phi_{St}^*(j\omega) \quad (15)$$

with the imitation transfer functions

$$\Phi_D^*(j\omega) = \bar{V}_0/\bar{V}_1 \quad (16)$$

and

$$\Phi_{St}^*(j\omega) = \bar{V}_{St}/\bar{V}_0 \quad (17)$$

respectively of the delay-line synthesizer and the stabilizing network. Virtually, these imitation transfer functions are here respectively the clamping transfer function of the delay-line synthesizer and the quotient of the clamping transfer function of the open loop (see Fig. 1b) and that of the delay-line synthesizer. If the exceptional behaviour of V_1 and V_2 is taken into account, eqns. (11)–(17) still hold for sinusoidal input V_1 . However, Φ_D^* and $\Phi_{circuit}^*$ are then, in general, not pure clamping transfer functions.

In general, $\Phi_{St}^*(j\omega)$ is not equal to the ordinary transfer function of the stabilizing network $\Phi_{St}(j\omega)$, for, in general, $V_{St}(k\tau)$ depends not only on $V_0(s\tau)$, with s the integers $s \leq k$, but also on the values of $V_0(t)$ for the times t in between.

We want to know how great can be the deviations of $\Phi_D^*(j\omega)$ and $\Phi_{St}^*(j\omega)$ from the transfer functions respectively of the system imitated, $\Phi_P(j\omega)$, and the stabilizing network, $\Phi_{St}(j\omega)$.

It can be shown [see eqn. (60)] that, neglecting the effect of the exceptional behaviour of V_1 and V_2 , for all integers s ,

$$\Phi_D^*(j\omega) = \Phi_D^*\left(j\omega + j\frac{2\pi s}{\tau}\right) \quad (18)$$

and

$$\Phi_{St}^*(j\omega) = \Phi_{St}^*\left(j\omega + j\frac{2\pi s}{\tau}\right) \quad (19)$$

Therefore it is quite clear that if one makes use of a discontinuous delay line one cannot, in general, expect to achieve a good imitation of an actual feedback system for angular frequencies other than those which are small in comparison with $2\pi/\tau$. In this respect the discontinuous delay-line synthesizer is no different from the continuous one.³

(3.2) Errors in Frequency Response of the Delay-Line Analogue

Apart from the inaccuracy in measuring the indicial response of the system and in observing the oscillograph, we distinguish four different causes of deviations between Φ_D^* and the transfer function of the imitated system Φ_P .

First, because of the finite length of the delay line, the indicial response of the system, $H(t)$, cannot, for times $t > T$ with

$$T = (N - 1\frac{1}{2})\tau + \sigma - t_0 \quad (20)$$

be imitated in detail. Instead of the actual indicial response,

[†] 'Clamping transfer functions' seems an adequate term for the transfer functions of clamping systems as defined in Reference 6.

$H(t)$, of the system, another indicial response, $H_C(t)$ is imitated, with, according to Section 3,

$$H_C(t) = H(t) \text{ for } t < T \quad (21)$$

For $t > T$ the difference $H(\infty) - H(T)$ is exponentially reduced to zero with time-constant T_T , which is given by

$$T_T = [H(\infty) - H(T)] / \left[\frac{dH(t)}{dt} \right]_{t=T} \quad (22)$$

The transfer function $\Phi_{P,T}$ corresponds to this new indicial response. The finiteness of T therefore causes a relative error in the frequency response of magnitude

$$\Delta_T = \frac{\Phi_{P,T} - \Phi_P}{\Phi_P} \quad (23)$$

For numerical evaluation of Δ_T , three different systems, P_4 , P_5 , and P'_4 have been chosen with the respective transfer functions:

$$\Phi_{P_4} = \frac{1}{(1 + j\omega T_0)^4} \quad (24)$$

$$\Phi_{P_5} = \frac{1}{(1 + j\omega T_0)^5} \quad (25)$$

$$\Phi_{P'_4} = \frac{1}{(1 + j\omega T_0)(1 + 2j\omega T_0)(1 + 4j\omega T_0)(1 + 8j\omega T_0)} \quad (26)$$

For each of these three systems the absolute value of Δ_T has been calculated in an elementary way and plotted in Fig. 6 as

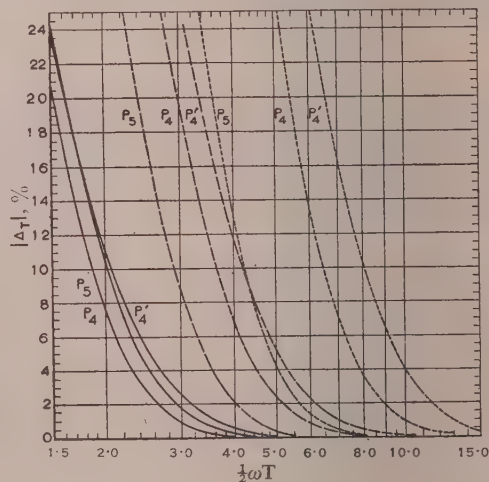


Fig. 6. $|\Delta_T|$.

— $\omega = \omega_{180}$
 --- $\omega = \omega_{230}$
 $\omega = \omega_{270}$

	P_4	P_5	P'_4
$\omega_{180}/\omega_{230}$	0.637	0.702	0.567
$\omega_{180}/\omega_{270}$	0.414	0.528	0.340

P_4 . Process with transfer function given by eqn. (24).

P_5 . Process with transfer function given by eqn. (25).

P'_4 . Process with transfer function given by eqn. (26).

a function of $\frac{1}{2}\omega T$ for those frequencies, ω_{180} , ω_{230} , and ω_{270} , for which its phase-lag is respectively 180° , 230° and 270° .

A second cause of deviation of Φ_D^* from Φ_P lies in the finiteness of the sampling period τ . To find its influence we suppose that the discontinuous delay line consists of an infinite number of sections, $N = \infty$, and that $V(k\tau) = H(k\tau - t_0)$ exactly, for an

input $V_1(t) = U(t - t_0)$, for all whole numbers k . It is shown in Section 6.3 that, if the clamping transfer function of this adjusted infinite delay-line synthesizer is Φ_{∞}^* , the relative deviation

$$\Delta_{\tau} = \frac{\Phi_{\infty}^* - \Phi_P}{\Phi_P} \quad (27)$$

would, for the above three systems, as for most higher-order systems, be quite adequately approximated by

$$\Delta_{\tau} = e^{-j\omega(t_0 + \frac{1}{2}\tau)} \left(\frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} - 1 \right) \quad (28)$$

Exactly the same formula would obtain if no integrator were applied, in contrast to the case with a continuous delay line.³ This is a consequence of the fact that the clamping transfer function of the discontinuous delay-line synthesizer, considered as a pure clamping system, is completely defined by the values of $V_0(k\tau)$ for an input $V_1(t) = U(t - t_0)$ (see Section 6.2).

Associated with this second cause of deviation of Φ_D^* from Φ_P is a third cause which arises from the practice of aiming, not at $V_0(k\tau) = H(k\tau - t_0)$, but, in accordance with the final part of Section 2.2, at

$$V_0(\frac{1}{2}k\tau + \sigma) = H(\frac{1}{2}k\tau + \sigma - t_0) \quad (29)$$

If eqn. (29) is exactly fulfilled for all whole numbers k (infinite delay line) and the time-constant of the integrator with which the delay-line synthesizer is provided is T_I , then, according to Section 6.3, the relative deviation of its clamping transfer function, $\Phi_{\infty, \sigma}^*$, from the transfer function of the imitated system, Φ_P ,

$$\Delta_{\tau, \sigma/\tau} = \frac{\Phi_{\infty, \sigma}^* - \Phi_P}{\Phi_P} \quad (30)$$

would for the above three systems, as for most higher-order systems, be quite adequately approximated by

$$\Delta_{\tau, \sigma/\tau} = e^{-j\omega(t_0 + \frac{1}{2}\tau)} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} [\rho e^{j\omega\sigma} + (1 - \rho)e^{j\omega(\sigma - \frac{1}{2}\tau)}] - 1 \quad (31)$$

$$\text{with} \quad \rho = \frac{1 - e^{-(\frac{1}{2}\tau - \sigma)/T_I}}{1 - e^{-\frac{1}{2}\tau/T_I}} \quad (32)$$

For $\tau/T_I < 1$, which, as will be seen in Section 4 and Fig. 15, is the case for the above three systems and probably for most higher-order systems, we can write

$$\Delta_{\tau, \sigma/\tau} \simeq e^{-j\omega(t_0 + \frac{1}{2}\tau)} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \left[\left(1 - \frac{2\sigma}{\tau} \right) e^{j\omega\sigma} + \frac{2\sigma}{\tau} e^{j\omega(\sigma - \frac{1}{2}\tau)} \right] - 1 \quad (33)$$

As is to be expected, $\Delta_{\tau, 0} = \Delta_{\tau}$.

The exceptional behaviour of V_1 and V_2 is a fourth possible source of error in the imitation of Φ_P . As already stated in Section 3.1, its effect is believed to be much smaller than that which would be caused by restriction $\alpha_1 = \alpha_2 = 0$. Now according to Section 6.3 this restriction would cause a relative deviation of Φ_D^* from Φ_P of

$$\Delta' = -j\omega\tau e^{j\omega t_0} \frac{\rho H(\sigma - t_0) + (1 - \rho)H(-\frac{1}{2}\tau + \sigma - t_0)}{\rho e^{j\omega\sigma} + (1 - \rho)e^{j\omega(\sigma - \frac{1}{2}\tau)}} / \Phi_P \quad (34)$$

and even this would be only a small contribution to the total relative deviation of Φ_D^* from Φ_P for the above three processes and probably for most higher-order processes (see Section 4 and compare Figs. 12 and 13). Therefore we neglect the effect of this fourth source of error.

From eqns. (28) and (33) we conclude that the best choice for t_0 is

$$t_0 = -\frac{1}{2}\tau \quad (35)$$

because for this value the absolute values of $\Delta_{\tau, 0}$ and $\Delta_{\tau, \frac{1}{2}}$ have rather sharp minima (as was stated in the last paragraph of Section 2.1, $\frac{1}{2}$ is the best value of σ/τ for the derivative action of the stabilizing network). For this value of t_0 the absolute value of $\Delta_{\tau, \sigma/\tau}$ is plotted in Fig. 7 as a function of $\frac{1}{2}\omega\tau$ for different values of σ/τ .

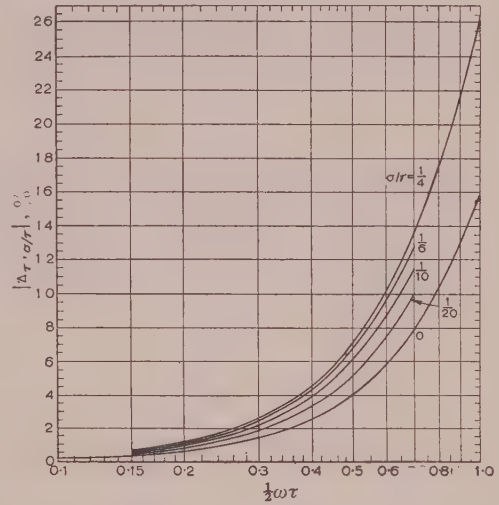


Fig. 7.— $|\Delta_{\tau, \sigma/\tau}| = |\Delta_{\tau, \frac{1}{2} - \sigma/\tau}|$.

Now one can expect that if Δ_{τ} and $\Delta_{\tau, \sigma/\tau}$ are small the relative deviation of the imitation transfer function of the delay-line synthesizer, Φ_D^* , from the transfer function of the imitated process, Φ_P ,

$$\Delta_{tot} = \frac{\Phi_D^* - \Phi_P}{\Phi_P} \quad (36)$$

is, in good approximation,

$$\Delta_{tot} = \Delta_{\tau} + \Delta_{\tau, \sigma/\tau} \quad (37)$$

For $t_0 = -\frac{1}{2}\tau$, we see that $\Delta_{\tau, \sigma/\tau}$ is real for $\sigma/\tau = 0$ or $\frac{1}{2}$ and nearly real for other values of σ/τ . Δ_{τ} is, in general, complex with an argument which seems to depend very much on ω and on T/T_0 . This is illustrated in the case of the imitation of H_B (see Fig. 4) by Figs. 7–10 of Reference 3. This dependence is quite different for different systems. For this reason we consider

$$\bar{\Delta}_{tot} = |\Delta_{\tau}| + |\Delta_{\tau, \sigma/\tau}| \quad (38)$$

instead of $|\Delta_{tot}| = |\Delta_{\tau} + \Delta_{\tau, \sigma/\tau}|$ as a good measure of the accuracy with which the discontinuous delay-line synthesizer can imitate a higher-order system.

(3.3) Errors in Frequency Response of the Stabilizing Network arising from the Behaviour of the Delay-Line Synthesizer

We restrict ourselves to the case that the stabilizing network has the properties of a 3-term controller and that its transfer function is

$$\Phi_{St}(j\omega) = P \left[1 + \frac{1}{j\omega T_i} + j\omega T_d \right] \quad (39)$$

It can then be considered as consisting of a proportional, an integral and a derivative element in parallel. The imitation

transfer function of the proportional element is exactly equal to its transfer function, because it simply multiplies the output for each time, and thus also for the sampling moments, by a constant factor P . The outputs of the integral and derivative elements at the sampling moments depend, not only on the output of the delay-line synthesizer at these moments, but also on its output during the times in between. The imitation transfer functions of the integral and derivative elements depend, therefore, on ω , τ , σ , and T_I .

The derivative element causes the most trouble, because the deviation of its imitation transfer function from its transfer function is, in general, much larger than the deviation of the imitation transfer function of the integral element from the latter's transfer function. In view of the derivative element, the best choice of σ is, according to the last paragraph of Section 2.2,

$$\sigma = \frac{1}{4}\tau \quad (40)$$

However, as stated in Section 2.1, the change-over time intervals of the different switches are, in general, neither constant nor equal. To find the order of magnitude of the influence of their variations on the imitation transfer function of the stabilizing network, we calculated the influence of a constant deviation of σ from $\frac{1}{4}\tau$. In Section 6.4 it is deduced that for small values of

$$\frac{1}{2}\delta = \frac{1}{2}\tau/T_I \quad (41)$$

and

$$e = \frac{\sigma - \frac{1}{4}\tau}{\frac{1}{4}\tau} \quad (42)$$

the imitation transfer function of the above stabilizing network is [eqns. (88), (89), (115) and (116)]

$$\Phi_{Si}^*(j\omega) = P \left[1 + \frac{1}{j\omega T_I} (1 + \Delta_i) + j\omega T_d (1 + \Delta_d) \right] \quad (43)$$

with

$$1 + \Delta_i = (1 + \Delta_{i,0})(1 - \delta\Delta_{i,\delta} - j\delta\bar{\Delta}_{i,\delta})(1 - e^2\Delta_{i,e} + je\bar{\Delta}_{i,e}) \quad (44)$$

$$1 + \Delta_d = (1 + \Delta_{d,0})(1 + \frac{1}{4}\delta e - \delta\Delta_{d,\delta})(1 - je\bar{\Delta}_{d,e}) \quad (45)$$

with the following functions of $x = \frac{1}{2}\omega\tau$:

$$\Delta_{i,0} = \frac{x}{\sin x} \frac{3 + \cos x}{4} - 1 \quad (46)$$

$$\Delta_{i,\delta} = \frac{1}{8} \left(\frac{1 - \cos x}{3 + \cos x} \right) \quad (47)$$

$$\bar{\Delta}_{i,\delta} = \frac{1}{24} \left(3 \tan x - \frac{\sin x}{3 + \cos x} \right) \quad (48)$$

$$\Delta_{i,e} = \frac{1 - \cos x}{3 + \cos x} \quad (49)$$

$$\bar{\Delta}_{i,e} = \tan \frac{1}{4}x - \frac{2 \sin x}{3 + \cos x} \quad (50)$$

$$\Delta_{d,0} = \frac{\tan \frac{1}{2}x}{\frac{1}{2}x} - 1 \quad (51)$$

$$\Delta_{d,\delta} = \frac{1}{8} \tan \frac{1}{2}x \quad (52)$$

$$\bar{\Delta}_{d,e} = \tan \frac{1}{2}x \quad (53)$$

Eqns. (46)–(53) are plotted in Figs. 8 and 9.

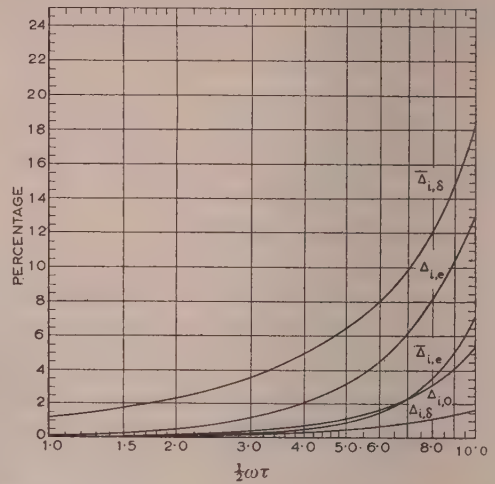


Fig. 8.— $\Delta_i \approx \Delta_{i,0} - \delta\Delta_{i,d} - e^2\Delta_{i,e} + j(e\bar{\Delta}_{i,e} - \delta\bar{\Delta}_{i,\delta})$.
 $\delta = \tau/T_I$ and $e = (\sigma - \frac{1}{4}\tau)/\frac{1}{4}\tau$.

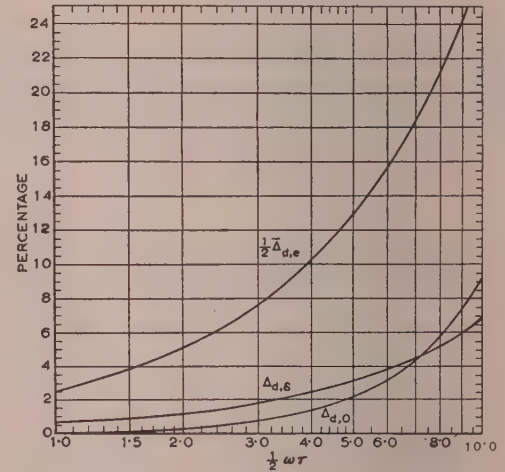


Fig. 9.— $\Delta_d \approx \Delta_{d,0} + \frac{1}{4}\delta e - \delta\Delta_{d,\delta} - je\bar{\Delta}_{d,e}$.
 $\delta = \tau/T_I$ and $e = (\sigma - \frac{1}{4}\tau)/\frac{1}{4}\tau$.

It follows from eqns. (44) and (45) that, for small values of eqns. (46)–(53),

$$\Delta_i \approx \Delta_{i,0} - \delta\Delta_{i,\delta} - e^2\Delta_{i,e} + j(e\bar{\Delta}_{i,e} - \delta\bar{\Delta}_{i,\delta}) \quad (54)$$

$$\Delta_d \approx \Delta_{d,0} + \frac{1}{4}\delta e - \delta\Delta_{d,\delta} - je\bar{\Delta}_{d,e} \quad (55)$$

From eqn. (55), the comparatively high values of $\bar{\Delta}_{d,e}$ according to Fig. 9, and the possible values of δ according to Fig. 15, it can be concluded that—as was already suggested in Section 2.2—if the stabilizing network is provided with a derivative action it is highly advantageous to keep σ as close as possible to $\frac{1}{4}\tau$.

(4) EVALUATION: ESTIMATE OF THE GREATEST ACCURACY AND THE PROCEDURE FOR ACHIEVING IT

In this Section we restrict ourselves again to the case in which the feedback system is a control loop consisting of a higher-order process and a 3-term controller.

It has already been discussed in Section 2.2 how best to imitate by means of a discontinuous delay-line synthesizer a higher-order

process of which the indicial response is known. It appeared from Section 3.2 that the best choice for t_0 is $t_0 = -\frac{1}{2}\tau$, and from the last paragraph of Section 3.3 that if a derivative controller action is active the best value for q is $\frac{1}{4}\tau$. There remains the choice of the time-scale reduction to be applied to the experimentally known indicial response of the process to adapt it to τ . To this end we shall discuss a method of finding the best value of $\frac{1}{2}\omega_{180}\tau$.

Since both $|\Delta_T|$ and $|\Delta_{\tau, \sigma/\tau}|$ are monotonic increasing functions of ω , we need consider them only for the upper boundary ω_{max} of the frequency range which is important. ω_{max} can, in general, be taken to be ω_{230} or ω_{180} , i.e. the frequency for which the phase lag of the process is 230° or 180° , according as the controller has or has not a derivative action. Therefore $|\Delta_T|$ was plotted in Fig. 6 for these frequencies. Also, curves were

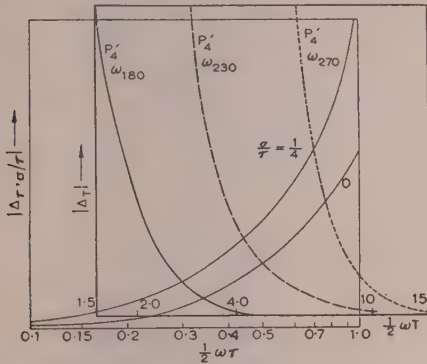


Fig. 10.—Method of finding the best value of $\frac{1}{2}\omega_{180}\tau$.

added for the frequency, ω_{270} , for which the phase lag of the three processes studied is 270° .

For given values of N and σ/τ and for each of the cases $\omega_{max} = \omega_{180}$, ω_{230} , and ω_{270} the best value of $\frac{1}{2}\omega_{180}\tau$ can now be found as follows. It is easily seen that with $t_0 = -\frac{1}{2}\tau$

$$T/\tau = N - 1 + \sigma/\tau. \quad (56)$$

We now lay a transparent copy of Fig. 6 on Fig. 7 with the abscissae superimposed in such a way that the coinciding values of $\frac{1}{2}\omega T$ and $\frac{1}{2}\omega\tau$ have the above ratio (see Fig. 10). This is possible because of their logarithmic scales. Both $|\Delta_{\tau, \sigma/\tau}|$ and $|\Delta_T|$ are now seen as functions of $\frac{1}{2}\omega\tau$ with $\omega = \omega_{180}$, ω_{230} , or ω_{270} . In all these three cases $|\Delta_T|$ is larger for P'_4 than for P_4 and P_5 for any value of $\frac{1}{2}\omega\tau$. Therefore we consider P'_4 as the standard process. In the case that P'_4 is imitated we consider as the best value of $\frac{1}{2}\omega_{max}\tau$ that value of $\frac{1}{2}\omega\tau$ for which the curve $|\Delta_T(\omega_{max})|$ for P'_4 and the curve $|\Delta_{\tau, \sigma/\tau}|$ for the given value of σ/τ intersect. The corresponding $\bar{\Delta}_{tot}(\omega_{max})$ is then the sum of the equal ordinates $|\Delta_T(\omega_{max})|$ and $|\Delta_{\tau, \sigma/\tau}|$ of this point of intersection. These values of $\frac{1}{2}\omega_{max}\tau$ and $\bar{\Delta}_{tot}(\omega_{max})$ do not differ much from those for which $\bar{\Delta}_{tot}(\omega_{max}) = |\Delta_T(\omega_{max})| + |\Delta_{\tau, \sigma/\tau}|$ for P'_4 is at its minimum. The former value of $\frac{1}{2}\omega_{max}\tau$ is smaller than the latter and therefore reduces the relative deviation of the imitation transfer function of the controller from its transfer function. From the above best value of $\frac{1}{2}\omega_{max}\tau$ the best value of $\frac{1}{2}\omega_{180}\tau$ for the case that P'_4 is imitated can be found simply by multiplying by the value of $\omega_{180}/\omega_{max}$ for P'_4 ($\omega_{180}/\omega_{230} = 0.568$; $\omega_{180}/\omega_{270} = 0.340$). We now take this value of $\frac{1}{2}\omega_{180}\tau$ to be, in general, the best. It is found that with this value both $|\Delta_T(\omega_{max})|$ and $|\Delta_{\tau, \sigma/\tau}(\omega_{max})|$ are smaller for P_4 and P_5 than for P'_4 .

In Fig. 11 the best values so defined of $\frac{1}{2}\omega_{180}\tau$ for $\sigma/\tau = \frac{1}{4}$ are

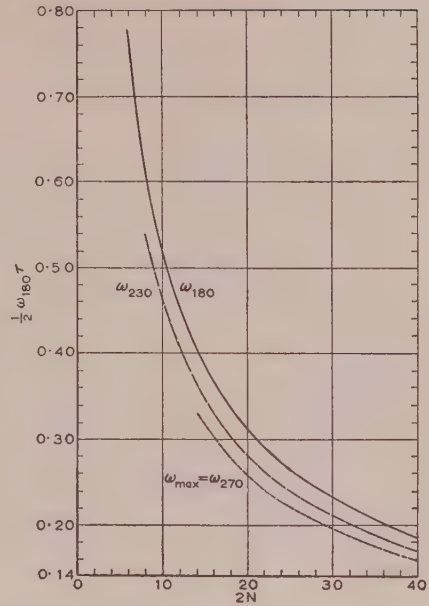


Fig. 11.—Value of $\frac{1}{2}\omega_{180}\tau$ for which, for process P'_4 , $|\Delta_T| = |\Delta_{\tau, \sigma/\tau}|$ at $\omega = \omega_{max}$.

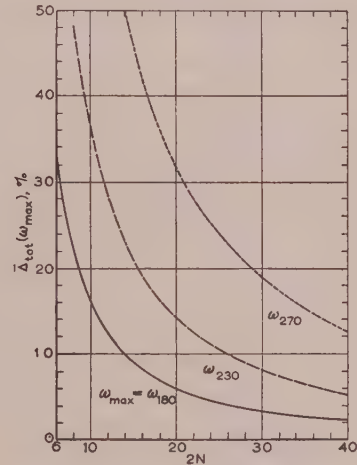


Fig. 12.—Value of $\bar{\Delta}_{tot}(\omega_{max})$ for process P'_4 if, at $\omega = \omega_{max}$, $|\Delta_T| = |\Delta_{\tau, \sigma/\tau}|$.

plotted as functions of the number of buffer amplifiers, $2N$, of which the delay line consists. The corresponding values of $\bar{\Delta}_{tot}(\omega_{max})$ for process P'_4 are plotted in Fig. 12.

Also, according to formula (34) $|\Delta'(\omega_{max})|$ with the above values of $\frac{1}{2}\omega_{180}\tau$ appears always to be larger for P'_4 than for P_4 and P_5 and is therefore plotted for P'_4 in Fig. 13. Comparison of Figs. 12 and 13 shows that, for P'_4 , $|\Delta'(\omega_{max})|$ is always much smaller than $\bar{\Delta}_{tot}(\omega_{max})$. It therefore seems indeed justifiable, as was assumed in Sections 3.1 and 3.2, to neglect the effect of the exceptional behaviour of V_1 and V_2 .

In view of the error in the frequency response of the controller [see formulae (54) and (55)] it is desirable to know about how large $\delta = \tau/T$ can be. To gain an idea of its possible values we proceeded as follows. Whereas according to Fig. 11 the

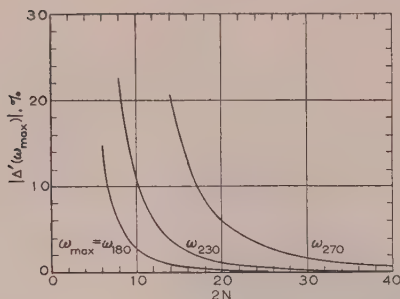


Fig. 13.— $|\Delta'(\omega_{\max})|$ if $|\Delta_T| = |\Delta_T|$ at $\omega = \omega_{\max}$ for process P_4 .

best values of $\omega_{180}\tau$ for $\sigma/\tau = \frac{1}{4}$ and for each of the three frequency ranges considered are monotonic decreasing functions of $2N$, the corresponding values of $\omega_{180}T$ are monotonic increasing functions of $2N$. For the relevant range of values of $\omega_{180}T$ we now calculate for each of the processes P_4 , P_5 , and P'_4 the appropriate values of $\omega_{180}T_I$. The results have been plotted in Fig. 14. It appears that the appropriate value of $\omega_{180}T_I$ for P_5 is for the whole range of values of $\omega_{180}T$ considered smaller than those for P_4 and P'_4 . It is now quite easy with the aid of the relations

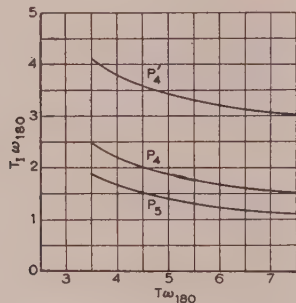


Fig. 14.— T_I .

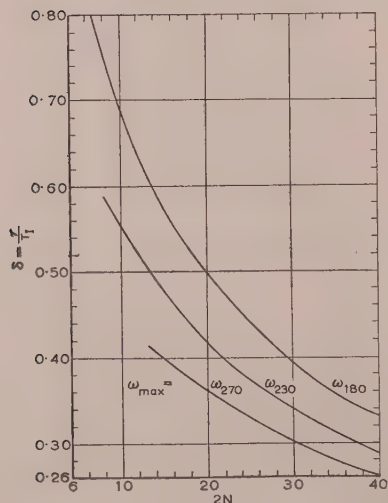


Fig. 15.— $\delta = \tau/T_I$ for process P_5 with reduced time scale.

illustrated in Figs. 11 and 14 to calculate $\omega_{180}T_I$ and $\delta = \tau/T_I$, the latter by simply dividing the values of $\omega_{180}\tau$ by the corresponding values of $\omega_{180}T_I$. The values of δ so calculated are plotted for P_5 in Fig. 15. From Fig. 14 it can be concluded that

the values of δ for P_4 and P'_4 are for the whole range of $2N$ considered respectively at least 1.23 and 1.73 times smaller than the corresponding values of δ for P_5 . From Fig. 15 it is seen that for a not too small number of buffer amplifiers, $2N$, $\delta < 1$, as was assumed in Section 3.2.

(5) REFERENCES

- (1) KALLMAN, H.: 'Transversal Filters', *Proceedings of the Institute of Radio Engineers*, 1940, 28, p. 302.
- (2) HEBB, M. H., HORTON, C. W., and JONES, F. B.: 'On the Design of Networks for Constant Time Delay', *Journal of Applied Physics*, 1949, 20, p. 616.
- (3) WESTCOTT, J. H.: 'The Continuous Delay-Line Synthesizer as a System Analogue' (see page 357).
- (4) JANSSEN, J. M. L.: 'Discontinuous Low-Frequency Delay Line with Continuously Variable Delay', *Nature*, 1952, 169, p. 148.
- (5) IDZERDA, H. H., ENSING, L., JANSSEN, J. M. L., and OFFEREINS, R. P.: 'Design and Application of an Electronic Simulator for Control Systems', *Transactions of the Society of Instrument Technology*, 1955, 7, p. 105.
- (6) JAMES, H. M., NICHOLS, N. B., and PHILLIPS, R. S.: 'Theory of Servomechanisms', (McGraw Hill, New York, 1947), p. 245.
- (7) MAGNUS, W., and OBERHETTINGER, H.: 'Formeln und Satze' (Springer, Berlin, 1948), Second edition, p. 217.

(6) APPENDICES

(6.1) Sinusoidal Input to Clamping System

If the output of a linear system $V_0(t)$ depends only on its input $V_1(t)$ at the sampling moments $t = k\tau$, it is easily seen that

$$V_0(k\tau) = \sum_{s=0}^{\infty} a_s V_1(k\tau - s\tau) \quad (57)$$

with the constants a_k .

If now the input is sinusoidal, or if at least

$$V_1(k\tau) = \mathcal{R}(\bar{V}_1 e^{j\omega k\tau}) \quad (58)$$

with constant complex value of \bar{V}_1 , it follows that

$$V_0(k\tau) = \mathcal{R}(\bar{V}_0 e^{j\omega k\tau}) \quad (59)$$

with

$$\bar{V}_0 = \bar{V}_1 \sum_{s=0}^{\infty} a_s e^{-j\omega s\tau} \quad (60)$$

(6.2) Clamping Transfer Function defined by Indicial Response at Sampling Moments

From eqn. (57) it follows that, for an input given by $V_1(t) = U(t - t_0)$,

$$V_0(k\tau) = \sum_{s=0}^k a_s \quad (61)$$

With the aid of these equations one can express the constants a_k in terms of $V_0(0)$ to $V_0(k\tau)$. \bar{V}_0/\bar{V}_1 is therefore, according to eqn. (60), completely defined by the values of $V_0(k\tau)$ for an input $V_1(t) = U(t - t_0)$.

(6.3) Imitation Transfer Function of a Discontinuous Delay-Line Synthesizer with an Infinite Number of Sections

We assume that the discontinuous delay-line synthesizer is a pure clamping system, which is exactly true only if

$$\alpha_1 = \alpha_2 = 0 \quad (62)$$

and that for an input

$$V_1(k\tau) = \begin{cases} 1 & \text{for } k \geq 0 \\ 0 & \text{for } k < 0 \end{cases} \quad (63)$$

the output is $V_0(t) = H_D(t - t_0)$ (64)

For an arbitrary input $V_1(t)$ the output at the sampling moments will then be

$$V_0(k\tau) = \sum_{s=-\infty}^{\infty} [V_1(k\tau - s\tau) - V_1(k\tau - s\tau - \tau)] H_D(s\tau - t_0) \quad (65)$$

If, now, the input is sinusoidal,

$$V_1(k\tau) = \mathcal{R}[\bar{V}_1 \varepsilon^{j\omega k\tau}] \quad (66)$$

it follows that $V_0(k\tau) = \mathcal{R}[\Phi^*(j\omega) \bar{V}_1 \varepsilon^{j\omega k\tau}]$ (67)

with

$$\Phi^*(j\omega) = \varepsilon^{-j\omega(\frac{1}{2}\tau + t_0)} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} j\omega\tau \sum_{s=-\infty}^{\infty} \varepsilon^{-j\omega(s\tau - t_0)} H_D(s\tau - t_0) \quad (68)$$

If now $H_D(k\tau - t_0) = \begin{cases} H(k\tau - t_0) & \text{for } k \neq 0 \\ 0 & \text{for } k = 0 \end{cases}$ (69)

$H_D(-t_0) = 0$ being a consequence of eqn. (62), it follows from eqns. (68) and (69) that

$$\begin{aligned} \Phi_{\infty}^*(j\omega) &= \varepsilon^{-j\omega(\frac{1}{2}\tau + t_0)} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \\ &\times \left[j\omega\tau \sum_{s=-\infty}^{\infty} \varepsilon^{-j\omega(s\tau - t_0)} H(s\tau - t_0) - j\omega\tau H(-t_0) \right] \quad (70) \end{aligned}$$

If, however, the discontinuous delay-line synthesizer is provided with an approximate integrator with time-constant T_I , and instead of eqn. (69) we have

$$H_D(\frac{1}{2}k\tau + \sigma - t_0) = \begin{cases} H(\frac{1}{2}k\tau + \sigma - t_0) & \text{for } k \neq -1, 0 \\ 0 & \text{for } k = -1, 0 \end{cases} \quad (71)$$

$H_D(-\frac{1}{2}\tau + \sigma - t_0) = H_D(\sigma - t_0) = 0$ being a consequence of eqn. (62), then

$$H_D(t - t_0) = 0 \quad \text{for } t < \sigma \quad (72)$$

$$H_D(t - t_0) = \frac{1 - \varepsilon^{-(t-\sigma)/T_I}}{1 - \varepsilon^{-\tau/2T_I}} H(\frac{1}{2}\tau + \sigma - t_0) \quad \text{for } \sigma < t < \sigma + \frac{1}{2}\tau \quad (73)$$

$$\begin{aligned} H_D(t - t_0) &= H\left(\frac{k}{2}\tau + \sigma - t_0\right) + \frac{1 - \varepsilon^{-(t - (k/2)\tau - \sigma)/T_I}}{1 - \varepsilon^{-\tau/2T_I}} \\ &\times \left[H\left(\frac{k+1}{2}\tau + \sigma - t_0\right) - H\left(\frac{k}{2}\tau + \sigma - t_0\right) \right] \end{aligned}$$

for $\frac{k}{2}\tau + \sigma < t < \frac{k+1}{2}\tau + \sigma$ $k \neq -1, 0$ (74)

and consequently $H_D(-t_0) = 0$ (75)

$$\begin{aligned} H_D(k\tau - t_0) &= \rho H(k\tau + \sigma - t_0) \\ &+ (1 - \rho) H(k\tau - \frac{1}{2}\tau + \sigma - t_0) \quad k \neq 0 \quad (76) \end{aligned}$$

with $\rho = [1 - \varepsilon^{-(\frac{1}{2}\tau - \sigma)/T_I}] / [1 - \varepsilon^{-\frac{1}{2}\tau/T_I}]$ (77)

It then follows from eqns. (68) and (75)–(77) that

$$\begin{aligned} \Phi_{\infty, \sigma}^*(j\omega) &= \varepsilon^{-j\omega(\frac{1}{2}\tau + t_0)} [\rho \varepsilon^{j\omega\sigma} + (1 - \rho) \varepsilon^{j\omega(\sigma - \frac{1}{2}\tau)}] \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \\ &\times \left\{ j\omega\tau \sum_{s=-\infty}^{\infty} \varepsilon^{-j\omega(s\tau + \sigma - t_0)} H(s\tau + \sigma - t_0) \right. \\ &\left. - j\omega\tau \frac{\varepsilon^{j\omega\tau_0} [\rho H(\sigma - t_0) + (1 - \rho) H(-\frac{1}{2}\tau + \sigma - t_0)]}{\rho \varepsilon^{j\omega\sigma} + (1 - \rho) \varepsilon^{j\omega(\sigma - \frac{1}{2}\tau)}} \right\} \quad (78) \end{aligned}$$

We now make use of the formula

$$\tau \sum_{s=-\infty}^{\infty} \varepsilon^{-j\omega(s\tau - \tau')} H(s\tau - \tau') = \sum_{s=-\infty}^{\infty} \frac{\exp \frac{-2\pi j s \tau'}{\tau} \Phi_P(j\omega + j \frac{2\pi s}{\tau})}{j(\omega + \frac{2\pi s}{\tau})} \quad (79)$$

which is equivalent to the well-known Poisson summation rule.⁷

It can be verified that, in general, for higher-order systems, for $\frac{1}{2}\omega\tau < 1$,

$$\begin{aligned} \left| j\omega \sum_{s=-\infty}^{\infty} \frac{\exp \frac{-2\pi j s \tau'}{\tau} \Phi_P(j\omega + j \frac{2\pi s}{\tau})}{j(\omega + \frac{2\pi s}{\tau})} - \Phi_P(j\omega) \right| / |\Phi_P(j\omega)| \\ \ll \left| \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} - 1 \right| \quad (80) \end{aligned}$$

if $\tau' = t_0$ or $t_0 - \sigma$.

Therefore it follows from eqns. (70), (78) and (79) that, to a good approximation,

$$\Phi_{\infty, \sigma}^*(j\omega) = \varepsilon^{-j\omega(\frac{1}{2}\tau + t_0)} \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} [\Phi_P(j\omega) - j\omega\tau H(-t_0)] \quad (81)$$

$$\begin{aligned} \Phi_{\infty}^*(j\omega) &= \varepsilon^{-j\omega(\frac{1}{2}\tau + t_0)} [\rho \varepsilon^{-j\omega\sigma} + (1 - \rho) \varepsilon^{j\omega(\sigma - \frac{1}{2}\tau)}] \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \\ &\times \left[\Phi_P(j\omega) - j\omega\tau \varepsilon^{j\omega t_0} \frac{\rho H(\sigma - t_0) + (1 - \rho) H(-\frac{1}{2}\tau + \sigma - t_0)}{\rho \varepsilon^{j\omega\sigma} + (1 - \rho) \varepsilon^{j\omega(\sigma - \frac{1}{2}\tau)}} \right] \quad (82) \end{aligned}$$

On the same lines it can be deduced that, if it were possible to make $H_D(k\tau - t_0) = H(k\tau - t_0)$ or $H_D(\frac{1}{2}k\tau + \sigma - t_0) = H(\frac{1}{2}k\tau + \sigma - t_0)$ for all integers k instead of eqn. (69) or (71), the delay-line synthesizer remaining a pure clamping system, the last factor in eqns. (81) or (82) would become simply $\Phi_P(j\omega)$. We thus find eqns. (28) and (31). Eqn. (34) is found by comparing eqn. (82) with its form changed in the way described above.

(6.4) The Imitation Transfer Function of the Stabilizing Network

If the output $V_0(t)$ of the discontinuous delay-line synthesizer, its integral

$$V_{int}(t) = \int_{-\infty}^{k\tau} V_0(t) dt \quad (83)$$

and its derivative $V_{der}(t) = \frac{dV(t)}{dt}$ (84)

at the sampling moments are respectively

$$V_0(k\tau) = \mathcal{R}(\bar{V}_0 \varepsilon^{j\omega k\tau}) \quad (85)$$

$$V_{int}(k\tau) = \mathcal{R}(\bar{V}_{int} \varepsilon^{j\omega k\tau}) \quad (86)$$

$$V_{der}(k\tau) = \mathcal{R}(\bar{V}_{der} \varepsilon^{j\omega k\tau}) \quad (87)$$

the imitation transfer function of the stabilizing network with transfer function (39) in series with this delay-line synthesizer is given by eqn. (43), with

$$1 + \Delta_i = j\omega \frac{\bar{V}_{int}}{\bar{V}_0} \quad (88)$$

$$1 + \Delta_d = \frac{1}{j\omega} \frac{\bar{V}_{der}}{\bar{V}_0} \quad (89)$$

In accordance with Section 3.1 we neglect the effect of the exceptional behaviour of V_1 and V_2 and assume that for an input as shown in eqn. (63) the output is given by eqn. (74) for all integers k . It follows then, that for an arbitrary input $V_1(t)$,

$$V_0(t) = \sum_{\substack{s=-\infty \\ (s-k \text{ even})}}^{+\infty} \left[V_1\left(\frac{k-s}{2}\tau\right) - V_1\left(\frac{k-s-2}{2}\tau\right) \right] F_s\left(t - \frac{k}{2}\tau - \sigma\right) \quad (90)$$

$$\text{for } \frac{k}{2}\tau + \sigma < t < \frac{k+1}{2}\tau + \sigma$$

with

$$F_s(x) = H\left(\frac{s}{2}\tau + \sigma - t_0\right) + \frac{1 - e^{-x/T}}{1 - e^{-\tau/2T_I}} \\ \times \left[H\left(\frac{s+1}{2}\tau + \sigma - t_0\right) - H\left(\frac{s}{2}\tau + \sigma - t_0\right) \right] \quad (91)$$

If now the input is sinusoidal [eqn. (58)] it follows that

$$V_0(t) = \mathcal{R} \left[\bar{V}_1 e^{j\omega(k/2)\tau} (1 - e^{-j\omega\tau}) \sum_{\substack{s=-\infty \\ (s-k \text{ even})}}^{+\infty} e^{-j\omega(s/2)\tau} F_s\left(t - \frac{k}{2}\tau - \sigma\right) \right] \quad (92)$$

$$\text{for } \frac{k}{2}\tau + \sigma < t < \frac{k+1}{2}\tau + \sigma$$

From eqn. (92) we find by integration by parts and summation of the results:

$$\int_{-\infty}^{k\tau} V_0(t) dt = \mathcal{R} \left\{ \bar{V}_1 e^{j\omega k\tau} (1 - e^{-j\omega\tau}) \right. \\ \left. \times \left[\frac{e^{-j\omega\tau}}{1 - e^{-j\omega\tau}} \sum_{s=-\infty}^{+\infty} e^{-j\omega s\tau} h_s + \sum_{s=-\infty}^{+\infty} e^{-j\omega s\tau} f_s\left(\frac{1}{2}\tau - \sigma\right) \right] \right\} \quad (93)$$

with

$$h_s = \frac{1}{2}\tau [H(s\tau - \frac{1}{2}\tau + \sigma - t_0) + H(s\tau + \sigma - t_0)] \\ + [H(s\tau + \frac{1}{2}\tau + \sigma - t_0) - H(s\tau - \frac{1}{2}\tau + \sigma - t_0)] \\ \times \frac{\frac{1}{2}\tau - T_I(1 - e^{-\tau/2T_I})}{1 - e^{-\tau/2T_I}} \quad (94)$$

$$f_s(x) = xH(s\tau - \frac{1}{2}\tau + \sigma - t_0) + [H(s\tau + \sigma - t_0) \\ - H(s\tau - \frac{1}{2}\tau + \sigma - t_0)] \frac{x - T_I(1 - e^{-x/T_I})}{1 - e^{-\tau/2T_I}} \quad (95)$$

From this we obtain, by means of eqn. (79),

$$\int_{-\infty}^{k\tau} V_0(t) dt = \mathcal{R} \left\{ \frac{1}{2} \bar{V}_1 e^{j\omega k\tau} \left\{ e^{-j\omega\tau} [R(\sigma - \frac{1}{2}\tau - t_0) + R(\sigma - t_0)] \right. \right. \\ + \psi(1)R(\sigma + \frac{1}{2}\tau - t_0) - \psi(1)R(\sigma - \frac{1}{2}\tau - t_0) \\ + (1 - e^{-j\omega\tau}) \left[\left(1 - \frac{2\sigma}{\tau}\right) R(\sigma - \frac{1}{2}\tau - t_0) \right. \\ \left. \left. + \psi\left(1 - \frac{2\sigma}{\tau}\right) R(\sigma - t_0) - \psi\left(1 - \frac{2\sigma}{\tau}\right) R(\sigma - \frac{1}{2}\tau - t_0) \right] \right\} \right\} \quad (96)$$

$$\text{with } R(C) = e^{j\omega C} \sum_{s=-\infty}^{+\infty} \frac{e^{j\omega s C/\tau} \Phi\left(j\omega + j\frac{2\pi s}{\tau}\right)}{j\left(\omega + \frac{2\pi s}{\tau}\right)} \quad (97)$$

$$\text{and } \psi(x) = \frac{x - \frac{2T_I}{\tau}(1 - e^{-x\tau/2T_I})}{1 - e^{-\tau/2T_I}} \quad (98)$$

On the other hand, one finds by differentiating eqn. (92)

$$\left[\frac{dV_0(t)}{dt} \right]_{t=k\tau} = \mathcal{R} \left\{ \bar{V}_1 e^{j\omega k\tau} (1 - e^{-j\omega\tau}) \frac{\frac{1}{\tau T_I} e^{-(\frac{1}{2}\tau - \sigma)/T_I}}{1 - e^{-\tau/2T_I}} \right. \\ \left. \times [R(\sigma - t_0) - R(\sigma - \frac{1}{2}\tau - t_0)] \right\} \quad (99)$$

For the three systems P_4 , P_5 and P'_4 and for most other higher-order systems one may neglect the differences between the factors

$$\sum_{s=-\infty}^{+\infty} \frac{e^{2\pi j s C/\tau} \Phi\left(j\omega + j\frac{2\pi s}{\tau}\right)}{j\left(\omega + \frac{2\pi s}{\tau}\right)} \quad (100)$$

of the different R 's [eqn. (97)] occurring in eqns. (96) and (99) in relation to the differences between their other factors $e^{j\omega C}$, and use the approximation

$$R(C) = \frac{1}{j\omega} e^{j\omega C} \Phi(j\omega) \quad (101)$$

We thus arrive at the formulae

$$\int_{-\infty}^{k\tau} V_0(t) dt = \mathcal{R} \left[\frac{1}{2} \bar{V}_1 e^{j\omega k\tau} \frac{\Phi(j\omega)}{j\omega} e^{j\omega(\sigma - \frac{1}{2}\tau - t_0)} \right. \\ \times \left\{ \psi\left(1 - \frac{2\sigma}{\tau}\right) e^{j\omega\tau} + \left[\psi(1) - \psi\left(1 - \frac{2\sigma}{\tau}\right) \right. \right. \\ \left. \left. + \left(1 - \frac{2\sigma}{\tau}\right) \right] + \left[1 - \psi\left(1 - \frac{2\sigma}{\tau}\right) \right] e^{-j\omega\tau} \right. \\ \left. \left. + \left[\frac{2\sigma}{\tau} - \psi(1) + \psi\left(1 - \frac{2\sigma}{\tau}\right) \right] e^{-j\omega\tau} \right\} \right] \quad (102)$$

and

$$\left[\frac{dV_0(t)}{dt} \right]_{t=k\tau} = \mathcal{R} \left[\bar{V}_1 e^{j\omega k\tau} \Phi(j\omega) j\omega e^{j\omega(\sigma - \frac{1}{2}\tau - t_0)} \right. \\ \left. \times \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} (1 - e^{-\frac{1}{2}j\omega\tau}) \frac{1}{j\omega} \frac{\frac{1}{T_I} e^{-(\frac{1}{2}\tau - \sigma)/T_I}}{1 - e^{-\tau/2T_I}} \right] \quad (103)$$

For $V_0(k\tau)$ we use eqns. (67) and (82) omitting in the latter the terms with H in accordance with the argument following it, and find with the aid of eqns. (83)–(89) that

$$1 + \Delta_i = \frac{1}{2} \left\{ \psi\left(1 - \frac{2\sigma}{\tau}\right) e^{j\omega\tau} + \left[\psi(1) - \psi\left(1 - \frac{2\sigma}{\tau}\right) \right. \right. \\ \left. \left. + \left(1 - \frac{2\sigma}{\tau}\right) \right] + \left[1 - \psi\left(1 - \frac{2\sigma}{\tau}\right) \right] \times e^{-j\omega\tau} \right. \\ \left. + \left[\frac{2\sigma}{\tau} - \psi(1) + \psi\left(1 - \frac{2\sigma}{\tau}\right) \right] e^{-j\omega\tau} \right\} / \left\{ \frac{\sin \frac{1}{2}\omega\tau}{\frac{1}{2}\omega\tau} \right. \\ \left. \times [\rho + (1 - \rho)e^{-\frac{1}{2}j\omega\tau}] \right\} \quad (104)$$

and
$$1 + \Delta_d = \frac{1 - e^{-\frac{1}{2}j\omega\tau}}{\rho + (1 - \rho)e^{-\frac{1}{2}j\omega\tau}} \frac{1}{j\omega} \frac{1}{T_I} \frac{e^{-(\frac{1}{2}\tau - \sigma)/T_I}}{1 - e^{-\tau/2T_I}} \quad (105)$$

with ψ and ρ defined by eqns. (98) and (77).

The most interesting case is $\sigma \simeq \frac{1}{4}\tau$. With the notations

$$e = (\sigma - \frac{1}{4}\tau)/\frac{1}{4}\tau \quad (106)$$

$$\delta = \tau/T_I \quad (107)$$

and
$$x = \frac{1}{2}\omega\tau \quad (108)$$

it follows that, for $\delta = 0$,

$$1 + \Delta_i = \frac{1}{4} \frac{x}{\sin x} \frac{3 + \cos x - 2je \sin x + e^2(\cos x - 1)}{1 - je \tan \frac{1}{2}x} \quad (109)$$

$$1 + \Delta_d = \frac{\tan \frac{1}{2}x}{\frac{1}{2}x} \frac{1}{1 - je \tan \frac{1}{2}x} \quad (110)$$

If $\delta < 1$ and e is small,

$$\psi(1) \simeq \frac{1}{2} [1 + \frac{1}{12}\delta] \quad (111)$$

$$\psi\left(1 - \frac{2\sigma}{\tau}\right) \simeq \frac{1}{8}(1 - e)^2(1 + \frac{1}{6}\delta) \quad (112)$$

$$\rho \simeq \frac{1}{2}(1 - e)(1 + \frac{1}{6}\delta) \quad (113)$$

$$\frac{1}{T_I} \frac{e^{-(\frac{1}{2}\tau - \sigma)/T_I}}{1 - e^{-\tau/2T_I}} \simeq \frac{2}{\tau} [1 + \frac{1}{4}\delta e - \frac{1}{6}\delta^2] \quad (114)$$

Making use of eqns. (104), (105) and (111)–(114) we find that, for small values of e and δ ,

$$\begin{aligned} 1 + \Delta_i = & \frac{x}{\sin x} \frac{3 + \cos x}{4} \left[1 + je \left(\tan \frac{1}{2}x - \frac{2 \sin x}{3 + \cos x} \right) \right. \\ & \left. - e^2 \frac{1 - \cos x}{3 + \cos x} \right] \times \left\{ 1 - \frac{1}{24}\delta \left[\frac{3 - 3 \cos x}{3 + \cos x} \right. \right. \\ & \left. \left. + j \left(3 \tan x - \frac{\sin x}{3 + \cos x} \right) \right] \right\} \quad (115) \end{aligned}$$

and

$$1 + \Delta_d = \frac{\tan \frac{1}{2}x}{\frac{1}{2}x} (1 - je \tan \frac{1}{2}x) [1 + \frac{1}{4}\delta(e - \frac{1}{2}j \tan \frac{1}{2}x)] \quad (116)$$

THE APPLICATION OF THE FREQUENCY-RESPONSE METHOD TO ELECTRICAL MACHINES

By S. K. SEN, B.E., and B. ADKINS, M.A., Member.

(The paper was first received 7th November, 1955, and in revised form 2nd February, 1956. It was published as an INSTITUTION MONOGRAPH in May, 1956.)

SUMMARY

The frequency-response method is well established in the analysis of control systems and networks, but has hitherto not been extensively used for analysing rotating machines. The method is of particular value when the magnetic circuit of the machine contains unlaminated iron in which eddy currents are set up when the flux changes. The paper describes a method of calculating the performance of the machine, and shows how the effect of the eddy currents can be taken into account.

Several methods of calculating the appropriate frequency-response curves have been worked out and experimentally verified. The application of the frequency-response curves to determine the performance characteristics is illustrated by calculations and tests for several simple transient conditions.

LIST OF SYMBOLS

- a = Half-thickness of solid iron section.
 f = Frequency.
 i = Instantaneous current.
 l = Mean length of the flux path in the solid iron.
 l_a = Axial length of solid iron section.
 l'_a = Effective axial length.
 p = d/dt = Heaviside operator.
 q, q' = Mesh lengths in relaxation pattern.
 r = Resistance.
 s = Slip.
 v = Instantaneous voltage.
 x = Co-ordinate across the solid iron section.
 x = Reactance (Section 5).
 B = Instantaneous flux density.
 F_0 = Residual in relaxation chart.
 H = Instantaneous magnetic force.
 N = Number of turns of the winding.
 Q = New variable replacing B in the relaxation chart.
 δ = Depth of penetration.
 $\theta = 2\pi ft$ = Angle in the time cycle.
 λ = Angle depending on the instant of switching.
 μ = Permeability.
 $\mu = \mu' - j\mu''$ = Complex permeability.
 ν = Instantaneous speed.
 ρ = Resistivity.
 τ = Torque in the turbo-alternator.
 ψ = Flux linkage (Section 5).
 $\omega = 2\pi f$ = Angular frequency.
 ω = Synchronous speed (Section 5).
 Φ = Flux.
 Ω = Variable in Floyd's method.
 1 = Heaviside unit-step function.
 $x_a(p), x_q(p)$ and $G(p)$ are operational functions depending on the winding impedances.

The following suffixes are used:

- a = Armature circuit.
 d = Direct axis.
 f = Field circuit.
 o = Value at the boundary.
 q = Quadrature axis.

Vectors are indicated by capital letters (except for certain Greek symbols) in bold face.

An asterisk with a symbol indicates the complex conjugate.

(1) INTRODUCTION

A frequency-response curve is a curve showing the relation between two quantities in a system, e.g. input voltage and output voltage, for operation over a range of frequencies. The ratio of the vectors corresponding to the two sinusoidally varying quantities at any frequency is determined as a complex number and plotted on an Argand diagram. The frequency-response curve is the locus of these points as the frequency varies from zero to infinity.

The method applies strictly to linear systems, but it can be used for an approximate study of a non-linear system if harmonics are neglected and the results are based on the fundamental values of the quantities. A rotating electrical machine often gives rise to more complicated problems than those encountered in linear control systems. The general equations of a machine are inherently non-linear, containing, for example, terms depending on the product of the speed and the currents. Further, saturation is always important to a greater or lesser degree. Some of the quantities in a machine can, however, be related by frequency-response curves, which can be used to determine the performance under various conditions of operation.

When the magnetic circuit of the machine contains solid unlaminated iron, the machine cannot be represented fully by simple circuits or by a simple set of equations. (The term 'solid iron' is used for iron which is not laminated.) It is possible, however, to determine the frequency-response curves, from which the performance characteristics can be deduced. The eddy currents, which flow in the solid iron, depend on saturation and hysteresis in the material. It is shown in the paper how these effects can be taken into account.

The investigation was made in three stages by considering successively

- A solid iron ring with a toroidal winding.
- A direct-current machine with a solid iron yoke.
- An alternator with a solid iron rotor.

(2) DETERMINATION OF THE FLUX AND INDUCED VOLTAGE IN A SOLID RING

(2.1) General

Before proceeding to the study of the d.c. machine, an investigation was made with a solid iron ring with a uniformly wound toroidal winding. The problem is to determine the voltage induced in the winding of the ring when a sinusoidal current of varying frequency flows in it. A thorough investigation of the ring, for which the magnetic circuit is a good deal simpler than

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Sen was formerly at Imperial College, and is now at the Bengal Engineering College, Howrah, India.

Mr. Adkins is in the Electrical Engineering Department, Imperial College of Science and Technology, University of London.

that of a machine, is recorded in Sections 2 and 3, and the results are extended in Section 4 to study the behaviour of a d.c. machine when alternating current flows in the field winding.

In an ideal inductance coil, having a magnetic system in which eddy currents, hysteresis and saturation do not occur, the locus of the current vector for an applied sinusoidal voltage of constant magnitude and variable frequency is a semicircle, as shown by the chain-dotted curve in Fig. 9. When the magnetic circuit is made of solid iron, either wholly, as in the ring, or partly, as in the d.c. machine, eddy currents flow in the iron and the current locus is distorted, as shown by the full-line curve in Fig. 9. The conditions are greatly affected by saturation and hysteresis, which distort the voltage wave and are important factors in determining the shape of the current locus.

The mild-steel ring is proportioned so that the axial length (12.9 cm) and the diameter (11.4 cm) are both much greater than the thickness (1.9 cm). The winding consists of 112 turns of wire of 0.122 cm diameter and has a resistance of 0.15 ohm (cold). A set of hysteresis loops of the iron, determined by direct-current tests using a search coil and a fluxmeter was used for making the calculations. The resistivity of the iron is 17×10^{-6} ohm-cm.

(2.2) Methods of Analysis for the Ring

At different values of the frequency f , the flux density varies across the thickness of the ring (along the x -axis) in the manner shown in Fig. 1. With direct current in the winding, the flux

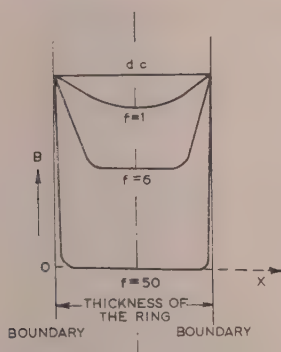


Fig. 1.—Variation of flux density across the ring at various frequencies.

density B is uniform, but with alternating current the amplitude falls off at the centre of the ring, because of eddy currents, by an amount which increases with frequency. The curve for 50 c/s shows clearly the so-called skin effect, which concentrates the flux at the two surfaces.

The ring can be analysed as a one-dimensional system by considering it to be equivalent to a bar, having infinite length in the circumferential direction and extending to infinity in the axial direction. Three different methods of analysis were used, each of which depends on a particular assumption made about the magnetic characteristics of the iron.

(a) *Constant-permeability method.*—The characteristic is assumed to be a straight line.

(b) *Relaxation method.*—The actual family of loop characteristics is used.

(c) *Complex-permeability method.*—The characteristics are assumed to be a family of ellipses.

In each method the manner of variation of B , both in space and time, can be determined for an alternating current of any mag-

nitude and frequency in the winding. The space variation across the thickness of the ring cannot be verified experimentally, but the time variation of the total flux can be checked by measuring the induced voltage.

The equation which determines the flux density is, in C.G.S. electromagnetic units,¹

$$\frac{\partial^2 H}{\partial x^2} = \frac{4\pi}{\rho} 10^{-9} \frac{\partial B}{\partial t} \quad (1)$$

The magnetic force, H_o , at the boundary is determined by the current flowing in the winding, and its value is used to obtain one of the boundary conditions required for the solution of the equation. If the current is sinusoidal, H_o also varies sinusoidally.

$$H_o = \frac{4\pi N}{10l} I_{max} \cos \omega t \quad (2)$$

(2.3) Method (a)

With $B = \mu H$, where the permeability μ is constant, eqn. (1) is linear, and consequently a sinusoidal current produces a sinusoidal flux. When transformed into a vector equation with frequency f , it becomes

$$\frac{d^2 H}{dx^2} = j \frac{4\pi\omega\mu}{\rho} 10^{-9} H \quad (3)$$

where H is a complex quantity representing the sinusoidally varying magnetic force.

Two alternative solutions are given below. The first solution applies generally and must be used at low frequencies. The second solution, which is much simpler, applies at high frequencies when the skin effect is so great that the flux density at the centre of the ring is negligible. The second boundary condition is then obtained by treating each half of the ring as a semi-infinite solid.

(2.3.1) Solution for Low Frequencies.

For a plate of thickness $2a$, the boundary conditions are as follows:

$$x = 0, H = H_o$$

$$x = a, \frac{dH}{dx} = 0$$

Solution of the equation gives the following expressions for the vectors representing the flux density and the induced voltage:

$$B = \frac{4\pi N\mu}{10l} I_{max} \frac{\cosh \alpha (x - a)}{\cosh \alpha a} \quad (4)$$

$$V = \frac{8\pi N^2 \mu j \omega l'_a}{\alpha l} I_{max} \tanh \alpha a \times 10^{-8} \quad (5)$$

$$\text{where } \alpha = \sqrt{\left(\frac{4\pi\mu\omega}{j\rho} \times 10^{-9} \right)} \quad (6)$$

(2.3.2) Solution for High Frequencies.

For a semi-infinite solid, the boundary conditions are

$$x = 0, H = H_o$$

$$x = \infty, H = 0$$

and the solution gives

$$B = \frac{4\pi N\mu}{10l} I_{max} e^{-\alpha x} \quad (7)$$

$$V = \frac{8\pi N^2 \mu j \omega l'_a}{\alpha l} I_{max} 10^{-8} \quad (8)$$

In applying the solution, the effective axial length must be assumed to be greater than the axial length of the ring, because the skin effect occurs all round the periphery. As an approximation the effective length l'_a is determined by the following equation:

$$l'_a = l_a + 2a - 2\delta \quad \dots \quad (9)$$

where δ is the depth of penetration, defined as the depth over which a uniform flux density equal to the surface density would produce the actual total flux. Its value, deduced for the semi-infinite solid, is

$$\delta = \sqrt{\left(\frac{\rho \times 10^9}{2\pi\mu\omega}\right)} \quad \dots \quad (10)$$

When δ exceeds a , l'_a is taken as equal to l_a .

(2.4) Method (b)

An accurate solution of eqn. (1), using the exact magnetic characteristics, can be obtained numerically by the method of relaxation.² The solution given by Butler and Sarma³ for thin laminations uses the relaxation method, but their assumptions about the distribution of H do not hold good for the present problem.

The solution is obtained using an angle $\theta = 2\pi ft$ to represent the time scale. If the meshes used for relaxation have dimensions

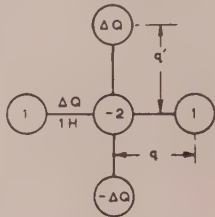


Fig. 2.—Relaxation pattern.

q along the x -axis and q' along the θ -axis, as indicated in Fig. 2, the finite-difference equation is

$$H_1 + H_3 - 2H_0 = \frac{Kq^2}{q'}(B_2 - B_4) \quad \dots \quad (11)$$

where $K = 8\pi^2 f \times 10^{-9}/\rho$.

For convenience in carrying out the relaxation, the variable B is replaced by a new variable Q , such that

$$Q = \frac{Kq^2}{q'}B \quad \dots \quad (12)$$

The residual equation is then

$$F_0 = H_1 + H_3 - 2H_0 - Q_2 + Q_4 \quad \dots \quad (13)$$

which leads to the relaxation pattern shown in Fig. 2.

The relation between H and Q at any mesh point is obtained from a different hysteresis loop for each value of x . It is, moreover, important to pass round the loop in the correct direction as time increases when moving along a vertical line corresponding to any value of x . The particular loop to be used for any point is not known at first, but it is readily determined in the course of the relaxation process. The solution obtained by method (a) was used as the starting-point of the relaxation. A typical relaxation chart, in which the upper left-hand figure at any node is Q , the lower left-hand figure is H and the right-hand figure is F_0 , is given in Fig. 3. The chart extends over half the thickness of the ring along the x -axis, and over half a period along the θ -axis.

θ	$x=0$	$x=0.01$	$x=0.02$	$x=0.03$	$x=0.04$	$x=0.05$	$x=0.06$	$x=0.07$	$x=0.08$	$x=0.09$	$x=0.10$
180°	-1.38 -9.56	-1.20 -8.25	0 -7.04	-0.03 -6.05	-0.08 -5.29	0.01 -4.82	-0.73 -4.66	0.01 -4.66	-0.66 -4.66	-0.04 -4.66	-0.02 -4.66
170°	-1.37 -9.40	-1.13 -7.84	0.01 -6.40	0 -5.31	0 -4.49	0 -3.97	0 -3.77	0 -3.77	0 -3.77	0 -3.77	0 -3.77
160°	-1.32 -8.98	-1.03 -7.03	-0.04 -5.54	-0.78 -4.33	-0.02 -3.46	-0.59 -2.95	-0.43 -2.78	-0.04 -2.78	-0.35 -2.78	-0.02 -2.78	-0.32 -2.78
150°	-1.19 -8.28	-0.91 -6.32	-0.06 -4.70	-0.64 -3.39	0 -2.50	0 -1.96	0 -1.78	0 -1.78	0 -1.78	0 -1.78	0 -1.78
140°	-1.05 -7.32	-0.75 -5.32	0.01 -3.68	-0.47 -2.42	0 -1.52	0 -0.98	0 -0.80	0 -0.80	0 -0.80	0 -0.80	0 -0.80
130°	-0.86 -6.14	-0.54 -4.07	0 -2.45	-0.26 -1.19	0.08 -0.29	-0.05 0.25	0.11 0.43	0.04 0.43	0.17 0.43	0.03 0.43	0.20 0.43
120°	-0.64 -4.78	-0.30 -2.65	0.04 -1.04	-0.03 0.14	0.01 0.99	0.16 1.50	0.08 1.67	0.30 1.67	0.04 1.67	0.36 1.67	0.06 1.67
110°	-0.40 -3.27	-0.06 -1.25	0 0.28	0.18 1.47	0.08 2.32	0.36 2.83	0.06 3.00	0.49 3.00	0.04 3.00	0.57 3.00	0.04 3.00
100°	-0.12 -1.66	0.19 0.24	0.03 1.64	0.39 2.79	0.06 3.54	0.56 3.97	0.02 4.12	0.68 4.12	0 4.12	0.74 4.12	0 4.12
90°	0.15 0	0.40 1.62	0.03 2.85	0.57 3.74	0.02 4.34	0.71 4.70	0 4.82	0.79 4.82	-0.02 4.82	0.83 4.82	-0.06 4.82
80°	0.40 1.66	0.61 3.07	0 4.07	0.75 4.74	0 5.19	0.85 5.46	0.01 5.56	0.90 5.56	-0.05 5.56	0.92 5.56	-0.03 5.56
70°	0.65 3.27	0.81 4.36	-0.01 5.06	0.90 5.50	0 5.78	0.94 5.96	-0.01 6.02	0.95 6.02	-0.01 6.02	0.97 6.02	-0.06 6.02
60°	0.87 4.78	0.99 5.59	-0.01 6.12	1.01 6.47	-0.02 6.70	1.00 6.85	-0.05 6.90	0.99 6.90	-0.02 6.90	0.99 6.90	-0.06 6.90
50°	1.06 6.14	1.10 6.56	0.01 6.77	1.06 6.88	0.01 6.98	1.01 7.04	0.04 7.06	1.01 7.06	0.02 7.06	1.01 7.06	0 7.06
40°	1.23 7.32	1.21 7.32	0 7.21	1.08 7.08	0.03 6.94	1.01 6.82	0.01 6.78	1.01 6.78	0.01 6.78	0.99 6.78	0.04 6.78
30°	1.33 8.28	1.21 7.97	0.03 7.66	1.11 7.17	0.05 6.93	1.03 6.72	0.06 6.67	1.00 6.67	0.07 6.67	0.97 6.67	0.08 6.67
20°	1.37 8.98	1.22 8.31	0.02 7.64	1.11 7.07	0.08 6.65	1.02 6.38	0.07 6.30	0.95 6.30	0.02 6.30	0.91 6.30	0 6.30
10°	1.38 9.40	1.23 8.46	0 7.54	1.09 6.75	0.03 6.14	0.95 5.78	0.02 5.65	0.87 5.65	0.03 5.65	0.78 5.65	0.02 5.65
0°	1.38 9.56	1.20 8.25	0 7.04	1.01 6.05	0.03 5.29	0.86 4.82	-0.01 4.66	0.73 4.66	0.01 4.66	0.66 4.66	0.04 4.66
BOUNDARY $x = a$											$x = 0$

Fig. 3.—Relaxation chart for 2 amp (r.m.s.) at 0.5 c/s.

Numbers at each node are: $\frac{Q}{H} | \frac{F_0}{H}$

The total flux at any instant is obtained by summation of the values of B along the x -axis, and the induced voltage is obtained from the total flux by numerical differentiation with respect to θ .

The accuracy of the relaxation method was checked by obtaining a solution based on a constant-permeability characteristic for comparison with the mathematical solution of method (a). The dotted lines in Fig. 6 were obtained by relaxation, and the points indicated by circles were calculated by method (a). The agreement was good, which showed that the choice of mesh size and the relaxation of the residuals were satisfactory for the purpose at hand.

(2.5) Method (c)

The method of calculation based on constant real permeability is relatively simple but is inaccurate because the assumptions about the magnetic characteristic are not correct. On the other hand, the relaxation method gives results which, as explained later, agree well with experiment, but require laborious numerical calculations. There is need for a reasonably simple method giving better accuracy than method (a). Method (c) is based on the concept of complex permeability.

The theory of complex permeability^{4,5} depends on the assumption that, if the magnetic force H at any point varies sinusoidally with time according to $H = H_{max} \cos \omega t$, the corresponding flux density B is given by an expression of the form

$$B = H_{max}(\mu' \cos \omega t + \mu'' \sin \omega t) \quad (14)$$

where μ' and μ'' are constants.

For a sinusoidal solution the relation between the vectors B and H is

$$B = \mu H \quad (15)$$

where $\mu = \mu' - j\mu''$ and is called the complex permeability. The assumption corresponds to a magnetic characteristic of elliptical shape.

With the above assumption, a sinusoidal applied current produces sinusoidal flux and voltage, and the solutions given in section 2.3 can be used if the complex permeability μ is inserted instead of the real permeability μ .

(2.6) Comparison of Measured and Calculated Results

(2.6.1) Unsaturated Conditions.

Tests were made on the ring to determine the induced voltage when a sinusoidal current flows in the winding. Oscillograms of voltage and current were recorded, and the instantaneous resistance drop was deducted from the terminal voltage in order to obtain the curve of induced voltage. The majority of the tests were made with a low current well below the point at which saturation commences. Fig. 4 shows the curves of induced voltage obtained by test and by the three methods of calculation, when a sinusoidal current of 2 amp (r.m.s.) at 50 c/s is passed through the winding. The instant of maximum current is taken as zero time.

As already mentioned, each of the methods of analysis depends on a particular assumption about the magnetic characteristics. Fig. 5 shows the three alternative curves for the maximum alternating magnetic force corresponding to 2 amp (r.m.s.).

Method (a).—The characteristic is the line (a), which is drawn along the mean line of the actual hysteresis loop.

Method (b).—The characteristics are the actual hysteresis loop (b) in Fig. 5, together with a family of smaller loops.

Method (c).—The characteristics are a family of ellipses, of which the largest, shown dotted in Fig. 5, is chosen to agree as closely as possible with the actual hysteresis loop. The method adopted is to make the ellipse pass through the point of maximum B and the point of zero H .

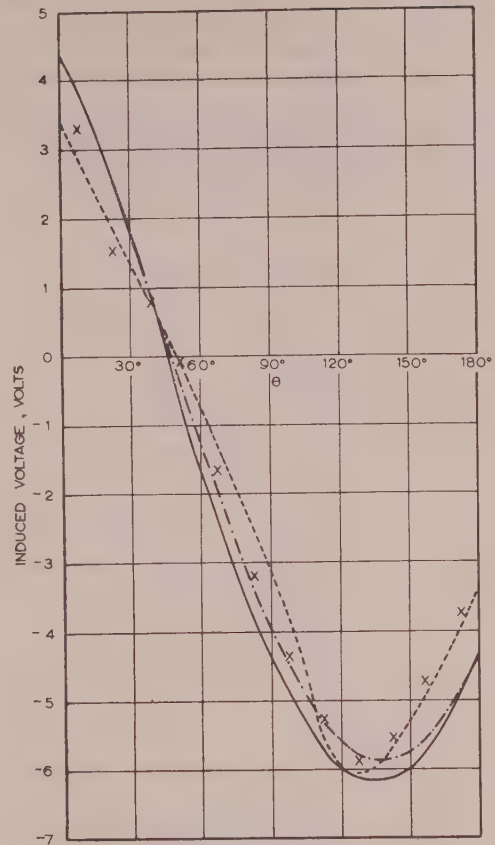


Fig. 4.—Induced voltage waveforms for 2 amp (r.m.s.) at 50 c/s.

— Test.
— Method (a).
x x x Method (b).
... Method (c).

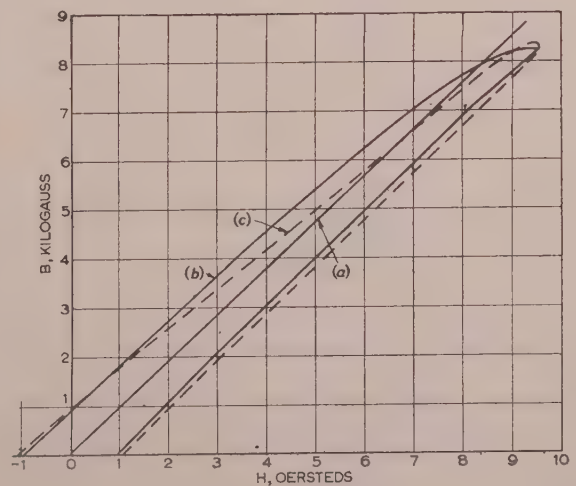


Fig. 5.—Magnetic characteristics associated with the three methods of calculation.

(a) Straight line, $\mu = 955$.
(b) Actual characteristic.
(c) Ellipse, $\mu = 863 - 94j$.

At the low current the saturation is negligible, but the hysteresis effect causes a distortion of the measured voltage wave. The curve calculated by relaxation shows the distortion, but methods (a) and (c) both give sinusoidal voltage waves. In order to obtain the frequency-response curves discussed in Section 3, the fundamental components of the voltage waves are used.

In order to show how the eddy currents affect the instantaneous distribution of flux density across the ring section, a series of curves is plotted in Fig. 6 for three instants during a half-cycle

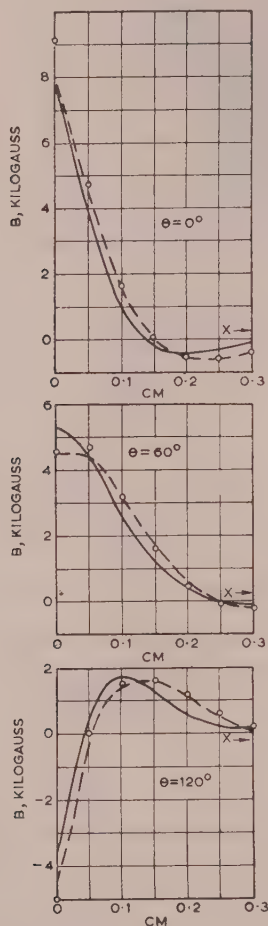


Fig. 6.—Distribution of flux density across the ring at different instants for 2 amp (r.m.s.) at 50 c/s.

○ ○ ○ ○ Method (a).
 ————— Method (b) with hysteresis.
 - - - - - Method (c) without hysteresis.

when the applied current is 2 amp (r.m.s.) at 50 c/s. The full-line curves show the flux density calculated by method (b), using the set of hysteresis loops, and the dotted curves show the flux density calculated by the same method but using the straight-line characteristic corresponding to constant permeability. The difference between the curves is evidently due to hysteresis, and they clearly show how the change of flux density at any point is delayed because of hysteresis. The points indicated by small circles were calculated by method (a), using the same value of μ as for the dotted line. The comparison between these points and the dotted line provides a check on the accuracy of the relaxation process, as mentioned in Section 2.4.

(2.7.2) Saturated Conditions.

Further tests and calculations were made with a larger current in the winding. Fig. 7 shows the curves of induced voltage when

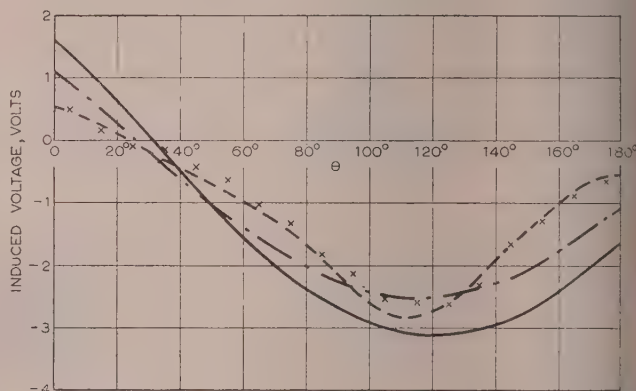


Fig. 7.—Induced voltage waveforms for 10 amp (r.m.s.) at 0.5 c/s.

— — — — — Test.
 Method (a).
 x x x x Method (b).
 - . - . - Method (c).

a sinusoidal current of 10 amp (r.m.s.) at 0.5 c/s flows in the winding, under which condition there is considerable saturation. The curve calculated by the relaxation method agrees well with the test curve, which is a distorted wave. The curves calculated by methods (a) and (c), however, differ greatly from that obtained by test. The effect of saturation is discussed further in Section 3.3.

(3) FREQUENCY-RESPONSE CURVES OF THE SOLID IRON RING

(3.1) Comparison of Measured and Calculated Results

If both the current and induced voltage are sinusoidal, the impedance of the coil, including its resistance, at any frequency can be determined as a complex number and plotted on an impedance locus curve. When the waveshape is distorted because of saturation and hysteresis the fundamental is used to determine the impedance, the value of which decreases as saturation increases. Fig. 8 shows the impedance loci determined by test and by the three methods discussed in Section 2 for a sinusoidal current of 2 amp (r.m.s.).

Several alternative methods of measuring the impedance at low frequencies down to 0.5 c/s were tried. The low-frequency current was taken either from an a.c. commutator machine running with a low slip, or from a self-excited cross-field generator (amplidyne) having a special electronic control system designed to produce a low-frequency supply. In either case a filter circuit was used to eliminate commutator and slot ripples.

The principal difficulty arose in measuring the phase angle, for which the following methods were tried:

- (i) Three-ammeter method and three-voltmeter method.
- (ii) Wattmeter method.
- (iii) Phase-shifter method. The Lissajous figure obtained by applying the voltage and current to a cathode-ray oscillograph was adjusted to indicate zero phase difference.
- (iv) Oscillograph method. Oscillograms of voltage and current for each condition were recorded with a Duddell oscillograph.

In order to obtain the best accuracy for the present investigation the last method was used, although the work involved in analysing the oscillograms was quite considerable. However, since the order of accuracy obtained in the calculations is

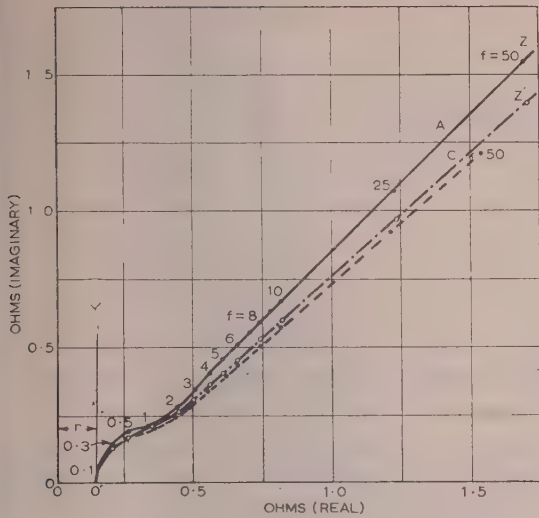


Fig. 8.—Impedance loci for 2 amp (r.m.s.).

----- Test.
 ——— Method (a).
 x x x x Method (b).
 - - - - Method (c).

relatively low, particularly when saturation occurs, it is probable that the values obtained by one of the other simpler methods would suffice for many practical purposes.

For a completely laminated ring with constant permeability the impedance locus at varying frequency would be the vertical straight line O'Y in Fig. 8, where OO' is the winding resistance r . For a solid ring with a pronounced skin effect, such that the theory of the semi-infinite solid applied, and having constant permeability, the impedance locus would be a straight line O'Z at 45° to the axis. The impedance locus determined by method (a) passes from one condition to the other. It starts at O' at zero frequency, is tangential to O'Y at first and then bends over to approach O'Z at higher frequencies.

Method (c) gives a similar result except that the curve approaches a line O'Z', where the angle between O'Z and the horizontal axis is less than 45° . The angle between O'Z and O'Z' represents a phase shift caused by hysteresis. It can be seen that the test curve agrees quite well with that obtained by method (c).

The effect of eddy currents and hysteresis shows up more clearly in the admittance-loci of Fig. 9, as explained in Section 2.1.

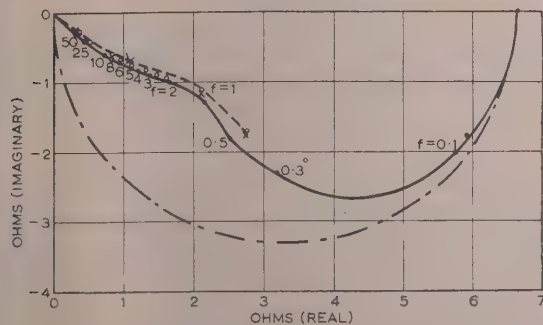


Fig. 9.—Admittance loci for 2 amp (r.m.s.).

----- Test.
 ——— Method (a).
 x x x x Method (b).
 o o o o Method (c).

The high-frequency part of the curve calculated by method (a) is initially at 45° to the real axis, but the angle of the measured curve and of those calculated by methods (b) and (c) is less because of hysteresis.

(3.2) Application of the Frequency-Response Curve to Determine the Response to a Suddenly Applied Voltage

With the solid ring the relation between the current in the coil and the applied voltage cannot be derived accurately as an operational formula, and a direct operational solution of a transient problem cannot be obtained. The admittance locus can, however, be calculated, and can be used to solve a transient problem by various known methods.

As an example, the admittance locus is used to determine the current that flows after a sudden application of voltage. In order to measure the transient current, an oscillograph shunt was connected in the circuit, increasing the total resistance to 0.213 ohm. Fig. 10 (full line) shows the modified admittance

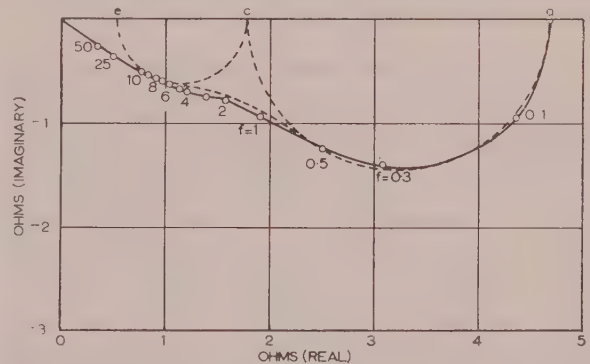


Fig. 10.—Admittance locus for 2 amp (r.m.s.) with $R = 0.213$ ohm.

——— Curve calculated by method (c).
 - - - - Approximate curve used to determine the transient current.

locus, calculated by method (c) for this resistance. The measured transient current is shown in Fig. 11 together with calculated curves obtained by two different methods.

The first method, described briefly in Appendix 8, is due to Floyd.⁶ It gives a good result but the numerical work is rather laborious.

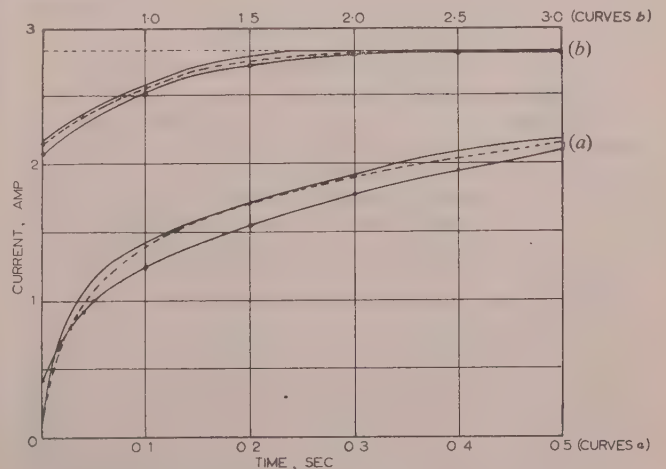


Fig. 11.—Transient current in the winding on the ring.

——— Test.
 - - - - Floyd's method.
 o o o o Algebraic method.

The second method is an algebraic one. The actual admittance is approximated by an expression of the form

$$Y(j\omega) = Y \frac{(1 + j\omega T_1)(1 + j\omega T_3)}{(1 + j\omega T_2)(1 + j\omega T_4)} \dots \quad (16)$$

where, T_1 , T_2 , T_3 and T_4 are constants. The corresponding admittance locus, shown dotted in Fig. 10, is a curve passing between two semicircles which have their centres on the real axis. The dotted curve is drawn for time-constants having the following values:

$$T_1 = 0.212 \text{ sec}, T_2 = 0.555, T_3 = 0.0098, T_4 = 0.0334$$

and is chosen to agree as nearly as possible with the actual locus. The transient response curve is then

$$i(t) = 2.83(1 - 0.646e^{-1.8t} - 0.240e^{-29.9t})$$

The current found by this algebraic method is shown in Fig. 11. It differs from the other curves because of the difference between the curves of Fig. 10 at the high-frequency end.

(3.3) The Effects of Eddy Currents, Hysteresis and Saturation

The analysis given above shows how the effects of eddy currents, hysteresis and saturation can, in general, be considered separately. Under conditions of low saturation the frequency-response curve can be satisfactorily calculated by means of a linear theory. As a rough approximation it may be stated that method (a), based on constant permeability, allows for eddy currents, while method (c), based on complex permeability, allows for both eddy currents and hysteresis. The effect of hysteresis appears separately as a charge in the angle of the impedance.

When saturation occurs it is necessary to select suitable values of the basic constants to fit the particular problem. The method, which is similar to that normally used to allow for saturation in machines, is an empirical one based on experience. In applying method (c) the value of μ' must be intermediate between that corresponding to maximum H and that at low saturation.

The purpose of the present paper is to establish a basic linear theory which allows for eddy currents and hysteresis in the solid iron. The application of the theory to the d.c. machine and the alternator in the later Sections is concerned only with conditions for which saturation is negligible.

(4) FREQUENCY-RESPONSE CURVES OF A D.C. MACHINE WITH A SOLID YOKE

(4.1) The Equivalent Ring with an Air Gap

The field winding of a d.c. machine, like the winding on the ring discussed above, is essentially an inductance coil, but it differs from the ring because the magnetic circuit is more complicated and the winding is concentrated in localized field coils. The methods already developed are modified and extended in this Section in order to study the effects of eddy currents and hysteresis in a d.c. machine when the yoke is made of solid iron. In addition to the voltage and current in the field winding, the voltage induced in the armature is also determined.

The investigation was made on a two-pole d.c. motor rated at 4.5 h.p., 200 volts, 1450 r.p.m., having the dimensions shown in Fig. 12. The poles are laminated but the yoke is solid. The calculations are made by first replacing the magnetic circuit by an equivalent iron ring with an air gap (thickness 3.5 cm, axial length 21 cm, mean circumferential length 24.2 cm, air-gap length 0.154 cm). The ring, which carries half the flux per pole of the machine and has the same cross-section as the yoke, is assumed to have no leakage. The mean circumference of the ring is equal to the mean length of the flux path per pole in the yoke. The

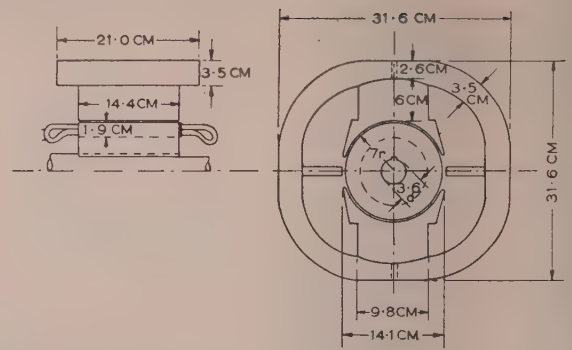


Fig. 12.—Dimensions of the d.c. machine. Mean air-gap length = 0.24 cm.

air gap of the ring is chosen so that it requires the same m.m.f. as the total m.m.f. per pole required by the air gap, armature core and teeth, and the laminated pole of the d.c. machine. The magnetic characteristics of the yoke material were determined by means of a search coil and a fluxmeter; saturation and hysteresis in the magnetic paths, other than the yoke, are neglected. The winding on the ring has the same number of turns as one pole of the field winding.

In the equivalent ring, since there is no leakage, the flux passing across the air gap is the same as that in the iron section. The flux density is assumed to be uniform at the air gap, and hence the m.m.f. across the air gap for a given flux can be calculated by using a constant factor. In the iron part, on the other hand, the flux is driven to the surface by the skin effect, and the m.m.f. required varies with frequency in both magnitude and phase. The m.m.f. can be calculated by the methods described in Section 2. The m.m.f. provided by the coil is equal to the vector sum of the two parts on the air gap and the iron.

The equivalence between the d.c. machine and the ring is based on the steady magnetization characteristic determined by calculation for the machine and verified by test. Fig. 13 shows saturation curves calculated by the normal design process of considering separately the five parts of the magnetic circuit. Of these, the air gap, the armature core and the armature teeth carry the air-gap flux, which determines the armature induced voltage, but the poles and the yoke carry a greater flux because of leakage. A leakage factor of 1.25 is assumed. Fig. 13 shows the total and partial m.m.f. for each value of open-circuit voltage, as indicated. The air-gap length in the ring is based on the straight line (c), allowing for the fact that the flux across the air gap of the ring is equal to the yoke flux of the machine.

The voltage induced in the field winding is calculated from the rate of change of the flux in the ring, which is assumed to be equal to that linking the actual field winding. The armature induced voltage, on the other hand, depends on the machine air-gap flux, which is equal to the ring flux divided by the leakage factor.

(4.2) Calculation of the Frequency-Response Curves

For the simplified magnetic circuit of the equivalent ring, the relation between the field current i_f , the field voltage v_f , and the armature voltage v_a can be calculated for alternating quantities at any frequency. Method (c), using complex permeability, was adopted as the main method of calculation, but a few curves were also determined by the relaxation method.

The differential eqn. (1) holds for the solid material as before, but the boundary conditions are different because of the air gap. The current in the winding determines the sum of the m.m.f. on

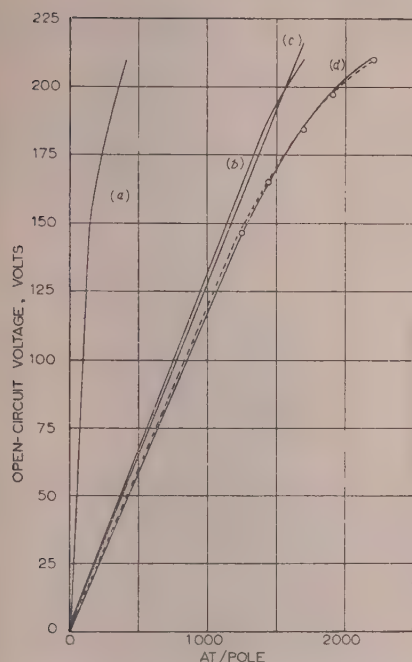


Fig. 13.—Saturation curves for the d.c. machine.

----- Test.
----- Calculated.
(a) Yoke only.
(b) Parts other than the yoke.
(c) Approximation to curve (b).
(d) Total.

the iron and the air gap, and the proportion of the two parts is not known until the flux has been determined. A solution by relaxation can only be obtained by trial and error, but the complex-permeability method requires only a simple calculation. The latter method is considered first.

4.2.1) Method (c)—Complex Permeability.

With the assumption that the hysteresis loops for the iron are ellipses, the flux and current are sinusoidal if the voltage is sinusoidal. With a given complex permeability μ , both the flux and current are proportional to the voltage and at a definite phase angle to it, at a given frequency. If the quantities are represented by vectors, the m.m.f., F_i , for the iron part of the circuit is related to the flux Φ by

$$F_i = K_i \Phi \quad (17)$$

where K_i is a complex number which varies with the frequency. The value of K_i can be calculated by the method of Section 2.5. The m.m.f., F_g , across the air gap of the ring is given by

$$F_g = K_g \Phi \quad (18)$$

where K_g is a real constant independent of the frequency.

The total m.m.f. provided by the current in the coil is the vector sum of F_i and F_g . The voltage across the coil is the sum of the voltage induced by Φ and the resistance-drop. In this way the impedance of the coil on the ring, and hence of the field winding of the d.c. machine, can be calculated at any frequency. From the flux Φ the armature voltage can also be calculated.

Thus the frequency-response curves for the field current and the armature voltage can be calculated for a given field voltage. A check should be made that the assumed value of μ is appropriate for the actual flux density, so that a correction can be

made if necessary. Figs. 14 and 15 show the curves calculated by method (c) compared with those obtained by test. The applied field voltage was sinusoidal at 25.7 volts (r.m.s.), which was low enough not to cause appreciable saturation under any condition.

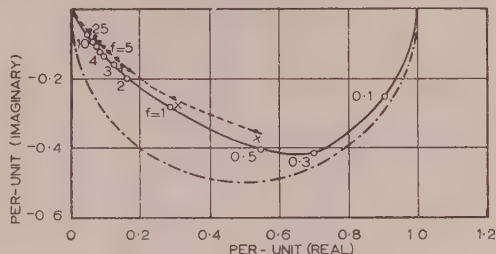


Fig. 14.—Loci of I_f/V_f for the d.c. machine.

----- Test.
x x x x Method (b).
----- Method (c).
----- Ideal locus.

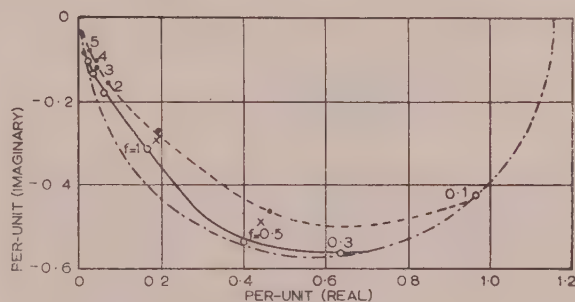


Fig. 15.—Loci of V_a/V_f for the d.c. machine.

----- Test.
x x x x Method (b).
----- Method (c).
----- Ideal locus.

(4.2.2) Method (b)—Relaxation.

In applying the relaxation method to the ring with an air gap, the procedure is similar to that used for the continuous ring considered in Section 2.4, but an additional process of trial and error is necessary because the values of H at the boundary are not known. A particular distribution of H must be assumed as a starting point, e.g. the sinusoidal distribution determined by method (c). The flux can be found as a function of time by relaxation, and hence the m.m.f. across the air gap, the current, and the resistance-drop in the coil. The voltage across the coil at any instant is then calculated, and if it does not agree with the actual applied voltage, the assumed boundary value of H must be corrected and the process repeated.

Fig. 16 shows the comparison between the waves of field current and armature voltage calculated by relaxation with those obtained by tests at a frequency of 0.5 c/s. The applied voltage is indicated as a sine wave having its maximum value at zero time.

The results show that the laborious relaxation method gives quite an accurate means of calculating the frequency-response curves, while the relatively simple complex-permeability method gives results sufficiently close for many practical purposes. Fig. 14 shows that the field current lags on the voltage by a small angle at very low frequencies, but as the skin effect increases, the magnitude of I_f decreases and the angle increases. At high frequencies the m.m.f.'s on the iron greatly exceeds that on the air gap, and the angle approaches a value just above 45° .

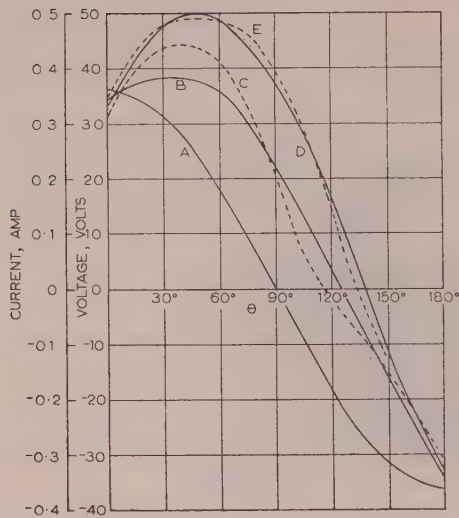


Fig. 16.—Waveforms of field current and armature voltage determined by test and by method (b) at 0.5 c/s.

- A. Applied field voltage.
- B. Field current by method (b).
- C. Field current by test.
- D. Armature voltage by method (b).
- E. Armature voltage by test.

The locus of the armature voltage (Fig. 15) is much nearer to the ideal semicircle than that of the field current (Fig. 14) because, to a first approximation, both V_g and V_f are induced by the same flux. The departure from the semicircle is mainly due to the manner in which the field current, and hence the resistance drop in the field winding, vary. Because of the assumption made about the leakage flux, the test curve, for both field current and armature voltage, is further from the semicircle than that obtained by either method of calculation.

(4.3) Application of the Frequency-Response Curve to Determine the Response to a Suddenly Applied Voltage

The transient build-up curve of the field current was calculated from the measured frequency-response curve of Fig. 14, using Floyd's method in the same way as for the solid iron ring considered in Section 3.2. The measured and calculated curves showing the field current after the winding is switched on to a constant voltage are shown in Fig. 17.

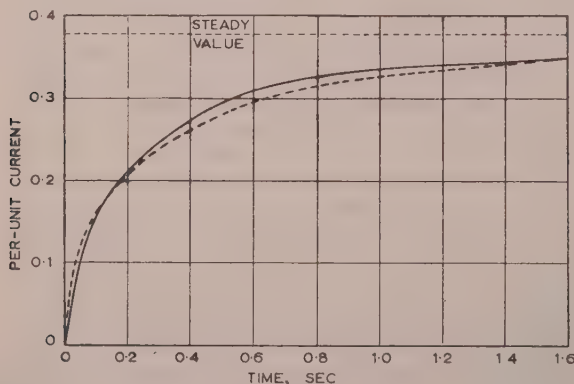


Fig. 17.—Transient field current of the d.c. machine.

- Test.
- Calculated by Floyd's method.

(5) THE FREQUENCY-RESPONSE CURVES OF A SYNCHRONOUS MACHINE WITH A SOLID ROTOR

(5.1) Operational Impedance

The two-axis theory of the synchronous machine, due to Park,⁷ is expressed by the following equations, in which all quantities are measured on a per-unit system. Some of the symbols and conventions differ from those used by Park, and follow those used in a paper by one of the authors.⁸

$$\left. \begin{aligned} v_d &= p\psi_d + v\psi_q + r_a i_d \\ v_q &= -v\psi_d + p\psi_q + r_a i_q \\ \omega\psi_d &= x_d(p)i_d + G(p)v_f \\ \omega\psi_q &= x_q(p)i_q \\ \tau &= \frac{\omega}{2}(\psi_q i_d - \psi_d i_q) \end{aligned} \right\} \quad (19)$$

The functions $x_d(p)$ and $x_q(p)$, which are known as the direct- and quadrature-axis operational impedances of the machine, are of particular importance in determining its performance. When the variables in the equations alternate at a frequency sf , they are converted into complex functions $x_d(js\omega)$ and $x_q(js\omega)$, which can be plotted on frequency-response curves. The investigation described in this Section is concerned primarily with these operational impedances, and deals with methods of measuring and calculating them.

The two-axis theory applies in the first instance to an 'ideal synchronous machine' having a fully-laminated magnetic circuit and field and damper windings consisting of coils wound symmetrically about the axes. If the damper winding is represented by one coil on each axis the operational impedances are given by the equivalent circuits shown in Fig. 18, where the various

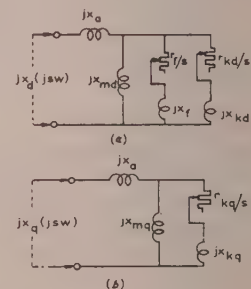


Fig. 18.—Equivalent circuits of a synchronous machine with a completely laminated magnetic circuit.

- (a) Direct axis.
- (b) Quadrature axis.

resistances and reactances are all constants. The corresponding loci of $x_d(js\omega)$ and $x_q(js\omega)$ are shown by the dotted lines in Fig. 19. For the quadrature axis the locus is a semicircle such that $x_q(js\omega)$ has the limiting values x_q (synchronous reactance) at $s = 0$ and x'_q (sub-transient reactance) at $s = \infty$. For the direct axis the locus passes between two semicircles which meet the real axis as shown at values x_d (synchronous reactance), x'_d (transient reactance) and x''_d (sub-transient reactance).

A synchronous machine, in which the rotor is made of solid iron—e.g. a normal turbo-alternator—is no longer ideal in the above sense. The eddy currents in the solid rotor provide a damping action, but the operational impedance loci, shown by full lines in Fig. 19, differ from the ideal curves. The effect is

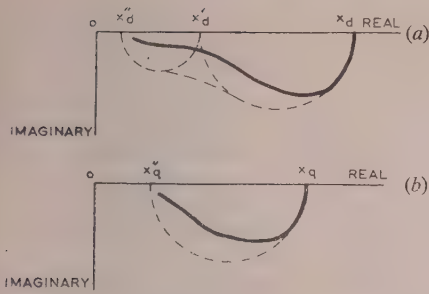


Fig. 19.—Frequency-response curves of a synchronous machine.

- (a) Locus of $x_d(j\omega)$.
 (b) Locus of $x_q(j\omega)$.
 --- Laminated machine.
 — Machine with solid rotor.

imilar to that obtained with the ring and the d.c. machine discussed earlier in the paper.

(5.2) Measurement of the Operational Impedances

The standstill impedance test on a synchronous machine with the rotor set on the direct or quadrature axis is a well-known method of measuring the direct- or quadrature-axis sub-transient reactance.⁹ If with the rotor in the direct-axis position a single-phase voltage V at the supply frequency is applied across two line terminals of the armature winding and a current I flows, the sub-transient reactance of a star-connected 3-phase machine is given approximately by

$$x'_d = \frac{V}{2I}$$

If the test is carried out at varying frequency the impedance does not vary in proportion to the frequency and is not a pure reactance because of the resistance of the armature winding and of the coupled field and armature windings. It approaches closer to the sub-transient reactance (at the appropriate frequency) as the frequency increases. For a machine with laminated poles, the value measured at normal frequency gives a reasonable approximation to the sub-transient reactance, but it is shown later than this is not true when the rotor is made of solid iron. Eqns. (19) show that, with $\nu = 0$ and $v_f = 0$,

$$v_d = \left[r_a + \frac{p}{\omega} x_d(p) \right] i_d$$

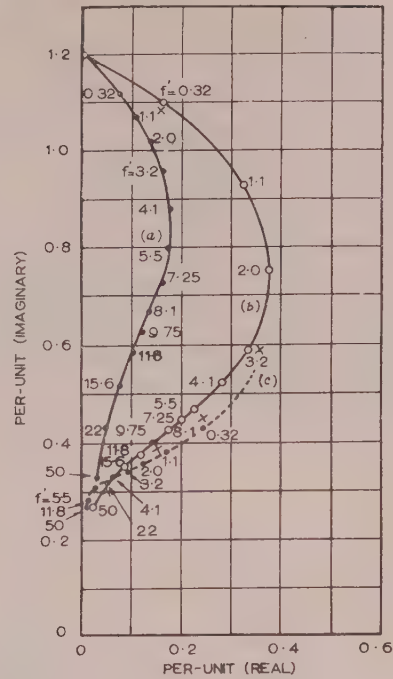
It can readily be shown from this result that the impedance measured on the direct-axis standstill impedance test at frequency f is

$$\left(\frac{V}{2I} \right)_d = r_a + j s x_d(j\omega) \quad . \quad . \quad . \quad (20)$$

imilarly, from the quadrature-axis test,

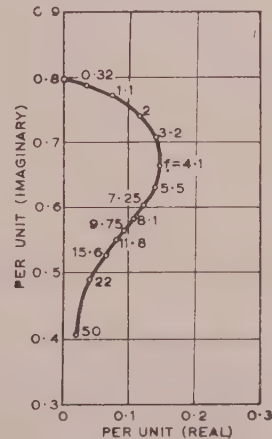
$$\left(\frac{V}{2I} \right)_q = r_a + j s x_q(j\omega) \quad . \quad . \quad . \quad (21)$$

The test therefore provides a means of obtaining by measurement the frequency-response curves giving the operational impedances $x_d(j\omega)$ and $x_q(j\omega)$ at varying frequencies. Tests of this kind were taken on a small turbo-alternator rated 440 volts, 3-phase, 50 c/s, 2-pole, 3 000 r.p.m. 125 kVA. In Fig. 20, which shows three curves for the direct axis, curve (a) was obtained with the field winding open-circuited, curve (c) with the field winding short-circuited, and curve (b) with the field circuit closed


 Fig. 20.—Experimental loci of $x_d(j\omega)$ for the turbo-alternator.

- (a) Field open-circuited.
 (b) Field closed through a resistance.
 (c) Field short-circuited.

× Calculated values for curve (b), derived from curve (a).


 Fig. 21.—Experimental locus of $x_q(j\omega)$ for the turbo-alternator.

through a resistance. Fig. 21 shows the curve for the quadrature axis.

The tests were taken with about one-quarter of the rated voltage, which corresponds to conditions of low saturation, except at the higher frequencies when the skin effect is considerable. As the waveform of the current was usually not sinusoidal, the fundamental was used in determining the impedance.

(5.3) Calculation of the Operational Impedances

The problem of calculating the impedance at any frequency was considered in relation to the direct-axis standstill test with the

field open-circuited. The impedance when the field circuit is closed can be deduced from the value when it is open by means of ordinary transformer theory, preferably using the equivalent circuit of a transformer with its secondary winding short-circuited. Hence, if a method can be developed for calculating curve (a), the other curves are readily deduced from it. A simple calculation, in which the transformer theory was used to derive curve (b) of Fig. 20 from curve (a), was made to verify this. The calculated points are shown by crosses in the Figure.

The condition of the standstill test with the field open is similar to that of the d.c. machine considered in Section 4, since the winding is essentially an inductance coil with a magnetic circuit having an air gap, a laminated part and a solid part, although the configuration of the magnetic circuit is even more complicated than that of the d.c. machine. The rather tentative method used to predetermine the frequency-response curve is based on the same method of replacing it by an equivalent ring with an air gap.

dimensions, as indicated in Fig. 23. The dimensions of the equivalent rectangle are taken to be AB in length and AC in thickness (see Fig. 22). The complete equivalent magnetic circuit consists of the solid rectangle in series with an air gap, and corresponds to the ring used for the d.c. machine, except that the proportions are quite different, since the length of the solid part is short compared with its width and thickness. Because of this and other factors the method is subject to considerable error. The length of the air gap was determined from the normal magnetic characteristics of the alternator by the same method as for the d.c. machine.

Table 1 shows the results of calculations made at a low frequency and at a high frequency. For a current of 42.5 amp (r.m.s.) the induced voltage, derived by deducting the stator resistance and leakage-reactance drops from the measured terminal voltage, is given in the first line. The phase of the induced voltage is taken as the reference phase, and the flux Φ

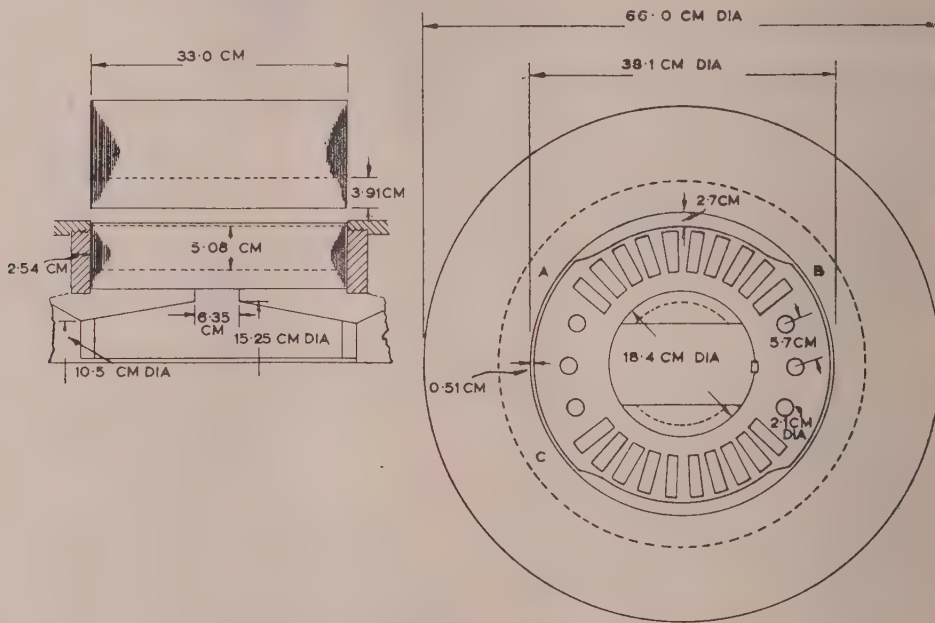


Fig. 22.—Dimensions of the turbo-alternator.

Fig. 22 shows sectional drawings of the stator and rotor cores of the turbo-alternator used for the investigation. It was unfortunate that the construction of this machine differed from that of a normal large turbo-alternator, in that the rotor was built with thick laminations mounted on a shaft instead of from solid material. The laminations were not insulated, and from the point of view of eddy currents, the conditions were intermediate between those in a fully-laminated rotor and a solid rotor. Moreover there were spaces in the magnetic circuit because the shaft was cut away to provide ventilation. The magnetic circuit was therefore more complicated and more difficult to analyse than that of a normal turbo-alternator, but the results of the investigation indicate that it would be possible to derive, on an empirical basis, a reasonable method of predetermining the frequency-response curves of a normal machine.

In the section of Fig. 22 the rotor direct-axis magnetic circuit passes horizontally with its upper and lower boundaries along the bottom surface of the slots. For the calculation the section is replaced by an equivalent rectangle of approximately the same

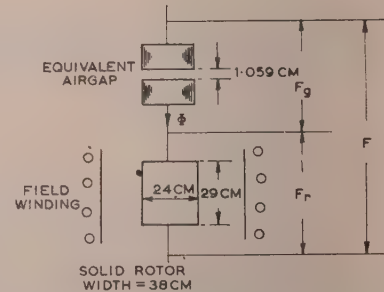


Fig. 23.—Equivalent magnetic circuit (direct axis) of the turbo-alternator.

therefore lags by 90° . The m.m.f. F_r on the solid part is calculated by the complex-permeability method, and is added to the m.m.f. F_g on the air gap. The total m.m.f. F , calculated in

Table 1

Frequency, c/s	0.32	50
Induced voltage, volts	$1.074 \cos \theta$	$26.1 \cos \theta$
Flux, Φ maxwells	$2.41 \times 10^6 \cos (\theta - 90^\circ)$	$0.374 \times 10^6 \cos (\theta - 90^\circ)$
M.M.F. F_g , AT	$932 \cos (\theta - 90^\circ)$	$144 \cos (\theta - 90^\circ)$
F_r	$463 \cos (\theta - 40.9^\circ)$	$1188 \cos (\theta - 35.3^\circ)$
Total m.m.f. F , AT	$1250 \cos (\theta - 74.5^\circ)$	$1290 \cos (\theta - 46.7^\circ)$
Measured m.m.f.	$1043 \cos (\theta - 85.5^\circ)$	$1043 \cos (\theta - 77.7^\circ)$

this way, is compared with the value obtained from the measured current.

Table 1 shows that the current passed by a given voltage is less than the calculated value and that the angle of lag is greater, indicating that the eddy-current effect is less powerful than that provided by a solid rotor. If, on the other hand, the calculation were based on the assumption that the outer laminated part of the rotor is not subject at all to eddy currents, the m.m.f. required would be much less, particularly at the high frequency. Thus in the machine tested the conditions are intermediate between those of a solid and a fully laminated rotor, and it is difficult to make an accurate calculation.

Calculations of the curve for the quadrature axis could be made by a similar method. For the machine represented in Fig. 22 the quadrature-axis flux would pass mainly through the laminated material, and if this part of the core were not subject at all to eddy currents, $x_q(j\omega)$ would be a constant, independent of frequency. The fact that $x_q(j\omega)$ decreases with the frequency, as shown in Fig. 21, shows that there are appreciable eddy currents in the thick uninsulated laminations.

(5.4) Application of the Frequency-Response Curves

In order to illustrate the use of the frequency-response curves of $x_d(j\omega)$ and $x_q(j\omega)$ to determine the performance of a synchronous machine, two important operating characteristics are considered.

- The torque/speed curve.
- The sudden short-circuit current.

In the following Sections the measured frequency-response curves of Figs. 20 and 21 are used to calculate these characteristics.

5.4.1 Calculation of the Torque/Speed Curve from the Frequency-Response Curves.

The ability of a synchronous machine to start from rest when connected to the supply depends on the torque produced at each value of speed between standstill and synchronism. The torque/speed curve shows the mean torque at any speed assuming that the speed is held constant. The torque actually pulsates at the curve shows the mean value.

The method of calculation is based on that derived by Linville¹⁰ from Park's eqns. (19). The axis voltages and currents pulsate at slip frequency and the following vector equations hold:

$$\left. \begin{aligned} V &= js\omega \Psi_d + (1-s)\omega \Psi_q + r_a I_d \\ jV &= -(1-s)\omega \Psi_d + js\omega \Psi_q + r_a I_q \\ \omega \Psi_d &= x_d(j\omega) I_d \\ \omega \Psi_q &= x_q(j\omega) I_q \\ \tau_m &= \frac{\omega}{2} \Re(\Psi_q^* I_d - \Psi_d^* I_q) \end{aligned} \right\} \quad (22)$$

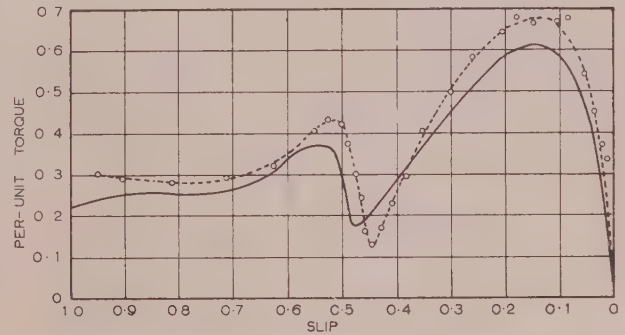


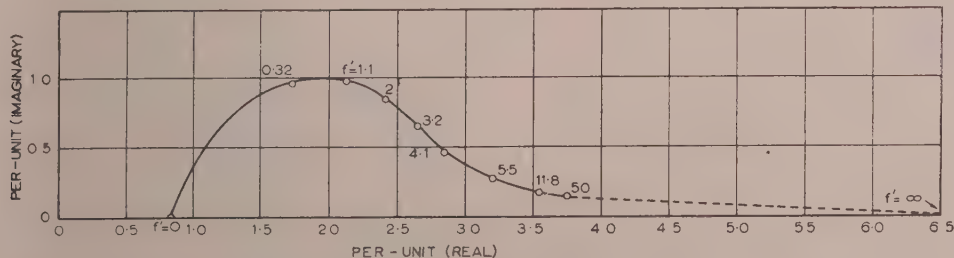
Fig. 24.—Torque/speed curve of the turbo-alternator.

----- Test.
————— Calculated.

The mean torque is given by the following expression, obtained by solving eqns. (22):

$$\tau_m = \frac{V^2}{2|D|^2} \Re \left[(A+B) + \frac{r_a}{1-2s} (2AB + |A|^2 + |B|^2) + r_a^2 (B|A|^2 + A|B|^2) \right] \quad (23)$$

where V is the supply voltage.


 Fig. 25.—Locus of $H(j\Omega)$ for the turbo-alternator.

$$A = \frac{1}{jx_d(js\omega)}$$

$$B = \frac{1}{jx_q(js\omega)}$$

$$D = 1 - \frac{s^2 r_a}{1 - 2s}(A + B) - \frac{r_a^2}{1 - 2s}AB$$

Fig. 24 shows the comparison between the calculated and measured curves for the 125 kVA turbo-alternator. The measured torque was obtained by coupling the alternator to a d.c. machine and measuring the d.c. output at a constant speed when the synchronous machine was operating as a motor with the field winding closed through a resistance. The calculated curve was obtained by using in eqn. (23) the values of the operational impedances given in Figs. 20 and 21. The values of $x_d(js\omega)$ were taken from curve (b) of Fig. 20, for which the field-circuit resistance was the same as for the torque/speed test.

The curves of Fig. 24 show clearly the dip in the torque in the neighbourhood of half-synchronous speed. The dip is more pronounced than in a normal synchronous motor with a squirrel-cage winding, because the damping effect of the rotor of the two-pole alternator is relatively weak on the quadrature axis.

(5.4.2) Calculation of the Short-Circuit Current from the Frequency-Response Curves.

The current that flows after a sudden symmetrical short-circuit on a 3-phase alternator can be split into an alternating component and a unidirectional component, each of which varies in magnitude as time passes. If the alternator is on open-circuit before the short-circuit is applied, the alternating component of the current in a phase winding is

$$i_a = i_d \cos(\omega t + \lambda)$$

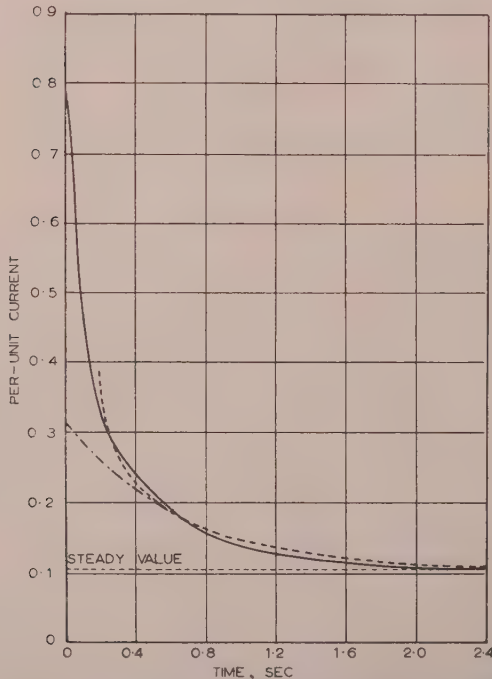


Fig. 26.—Alternating component of the short-circuit current.

— Test.
 - - - Transient component deduced from the test.
 Calculated.

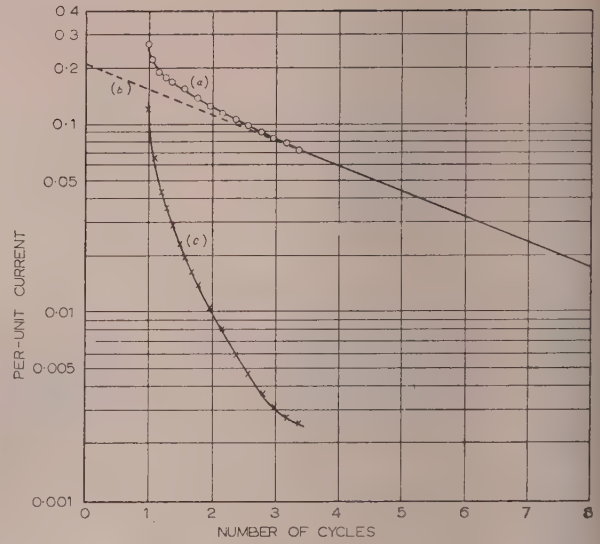


Fig. 27.—Semi-logarithmic plot of the short-circuit current.

(a) Alternating component minus the steady value.
 (b) Transient component.
 (c) Sub-transient component.

where i_d , the direct-axis current, is given by the following operational equation, which is derived from Park's equations:⁸

$$i_d = \frac{1}{x_d(p)} V1 = H(p)1$$

where V is the peak voltage before short-circuit.

The locus of $H(j\Omega)$, derived from the locus of $x_d(js\omega)$ in Fig. 20, is plotted in Fig. 25. The function does not tend to zero as Ω tends to infinity and Floyd's method does not apply directly. However, if the function is written as follows:

$$H(j\Omega) = H(\infty) + H'(j\Omega)$$

Floyd's method can be used for the second term $H'(j\Omega)$, which is given in Fig. 25 if the origin is transferred. The calculated curve of the alternating component of the short-circuit current is compared with the measured curve in Fig. 26.

The short-circuit current normally depends on the transient and sub-transient reactances and time-constants. In order to determine the constants from the short-circuit test the difference between the measured current and the steady current is plotted against time on semi-logarithm paper in Fig. 27 [curve (a)]. The dotted line is a straight line which coincides with the lower part of the measured curve and is used to obtain the transient reactance and the transient time-constant. The difference between curve (b) and curve (a) is replotted as curve (c). For an alternator with a laminated rotor, curve (c) would be a straight line and would determine the sub-transient reactance and the sub-transient time-constant. However, when the rotor is made of solid iron this is by no means the case. Thus the concept of sub-transient reactance does not strictly apply unless a different value is used for every different condition.

For the frequency-response curve of Fig. 25 no measurements were made at frequencies above 50 c/s. At infinite frequency no flux can penetrate into the solid rotor, and consequently the corresponding sub-transient reactance must be equal to the armature leakage reactance. The leakage reactance could not be determined directly, but it is known that the true leakage reactance is less than the Potier reactance, which was found to be

191 per unit. The leakage reactance was taken to be 0.154, giving a value for $H(\infty)$ of 6.5, as shown on Fig. 25.

The initial part of the curve of short-circuit current depends mainly on the high-frequency end of the frequency locus. Fig. 25 shows that, for determining the initial peak of short-circuit current, the sub-transient reactance must be taken to be much less than the value measured on a 50 c/s standstill impedance test. The appropriate value is approximately the leakage reactance.

(6) CONCLUSION

The frequency-response curves of the d.c. machine and the alternator provide a valuable method of studying their performance, particularly when the magnetic circuit contains solid iron in which eddy currents flow. The paper describes some methods of calculating the curves, and discusses some applications to determine the performance of the machines.

Although the accuracy of the calculations is not high, they show clearly the way in which the operation of the machine is affected by the eddy currents in the solid iron. Further work is needed to derive a reliable method of calculating the impedance loci of an alternator with a solid rotor, and for allowing empirically for the effect of saturation.

(7) ACKNOWLEDGMENT

The authors wish to thank the British Thomson-Houston Co., Ltd., for information and assistance in carrying out the investigation.

(8) REFERENCES

- (1) THOMSON, J. J.: 'On the Heat produced by Eddy Currents in an Iron Plate Exposed to an Alternating Magnetic Field', *Electrician*, 1892, **28**, p. 599.
- (2) SOUTHWELL, R. V.: 'Relaxation Methods in Theoretical Physics' (Oxford University Press, 1946).
- (3) BUTLER, O. I., and SARMA, M. R.: 'Relaxation Methods applied to the Problem of A.C. Magnetization of Ferromagnetic Laminæ', *Proceedings I.E.E.*, Paper No. 1026 M, January, 1951 (**98**, Part II, p. 389).
- (4) PRESS, A.: 'Treatment of Harmonics in A.C. Theory by means of Harmonic Algebra', University of California publication in Engineering, 1919, **2**, p. 141.
- (5) MACFADYEN, K. A.: 'Vector Permeability', *Journal I.E.E.*, 1947, **94**, Part III, p. 407.
- (6) BROWN, G. S., and CAMPBELL, D. P.: 'Principles of Servomechanisms' (John Wiley, 1948), p. 332.
- (7) PARK, H.: 'Two-reaction Theory of Synchronous Machines', *Transactions of the American I.E.E.*, 1929, **48**, p. 716.
- (8) ADKINS, B.: 'Transient Theory of Synchronous Generators connected to Power Systems', *Proceedings I.E.E.*, Paper No. 1051 S, November, 1950 (**98**, Part II, p. 510).
- (9) WRIGHT, S. H.: 'Determination of the Synchronous Machine Constants by Test', *Transactions of the American I.E.E.*, 1931, **50**, p. 1336.

- (10) LINVILLE, T. M.: 'Starting Performance of Salient-Pole Synchronous Motors', *ibid.*, 1930, **49**, p. 531.

(9) APPENDIX

Floyd's Method of Determining the Transient Response from the Frequency Response⁶

If $H(j\Omega)$ is the transfer function of the system, the following conditions must be complied with:

- (a) $H(j\Omega)$ is a function which can be expressed as the ratio of two polynomials in j with real constant coefficients.
- (b) $H(j\Omega)$ has no poles on the right-hand half of the j -plane or on the imaginary axis.
- (c) Limit $\Omega \rightarrow \infty H(j\Omega) = 0$.

Conditions (a) and (b) are complied with for electrical systems of the type considered, because the solid magnetic material can be replaced by a network having a large number of elements with self and mutual inductance.

According to Floyd's method, the transient response to a unit impulse is given approximately by

$$h(t) = \frac{2}{\pi} \int_0^{\infty} \Re[H(j\Omega) \cos(\Omega t)] d\Omega$$

The integration is performed by plotting the real part of $H(j\Omega)$

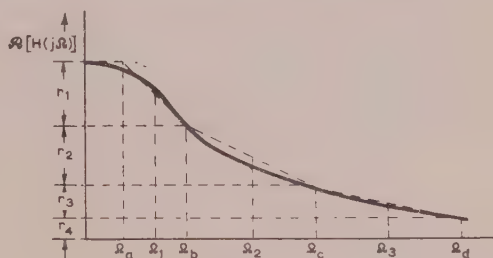


Fig. 28.—Approximation to the $\Re[H(j\Omega)]/\Omega$ curve.

against Ω , as in Fig. 28, and drawing a series of straight lines which approximate to the curve.

$$\text{Then } h(t) = \sum_{n=1}^k \frac{2}{\pi} A_n \left(\frac{\sin \Omega_n t}{\Omega_n t} \right) \left(\frac{\sin \Delta \Omega_n t}{\Delta \Omega_n t} \right)$$

$$\begin{aligned} \text{where } \Omega_1 &= \frac{1}{2}(\Omega_a + \Omega_b); \Omega_2 = \frac{1}{2}(\Omega_b + \Omega_c) \\ \Delta \Omega_1 &= \frac{1}{2}(\Omega_b - \Omega_a); \Delta \Omega_2 = \frac{1}{2}(\Omega_c - \Omega_b) \\ A_1 &= r_1 \Omega_1; A_2 = r_2 \Omega_2 \end{aligned}$$

The response to a unit step function is found by integrating $h(t)$ with respect to t .

MICROWAVE MEASUREMENTS WITH A LOSSY VARIABLE TERMINATION

By H. M. ALTSCHULER, M.E.E., and Prof. A. A. OLINER, Ph.D.

(The paper was first received 2nd December, 1955. It was published as an INSTITUTION MONOGRAPH in May, 1956.)

SUMMARY

The use of a lossy variable termination becomes an undesirable necessity in certain measurement situations, while in others its deliberate use facilitates the measurement of low-loss structures. In both of these cases, the present approach permits the analysis of the data to proceed in a manner identical to that employed when lossless short-circuits are used, i.e. with already familiar procedures. It results in a composite 4-terminal network from which the desired parameters are separated using particularly simple relations.

Applications are made to a variety of measurements, including those of dissipative, lossless and symmetric 4-terminal networks, attenuation and attenuation constants.

LIST OF SYMBOLS

- A_D = Dissipative attenuation of 4-terminal network, dB.
 A_R = Reflective attenuation of 4-terminal network, dB.
 A_T = Total attenuation of 4-terminal network, dB.
 d = Length of sample waveguide in attenuation-constant measurement.
 D = Distance from an arbitrary zero to the voltage minimum in the slotted line (measured from the arbitrary zero to the left).
 D_R = Value of D measured when the slotted line is short-circuited.
 D_0 = A parameter of the tangent relation network of Fig. 7.
 n = Transformer turns ratio.
 r_s = Voltage standing-wave ratio (v.s.w.r.) corresponding to the reflection coefficient $|\Gamma_s|$.
 r_{max} = Maximum v.s.w.r.
 r_{min} = Minimum v.s.w.r.
 S = Distance from an arbitrary zero to a voltage minimum in the variable termination (measured from the arbitrary zero to the right).
 S_R = Value of S when a voltage minimum is located at the input terminal plane of the variable termination (or short-circuit).
 S_0 = A parameter of the tangent relation network of Fig. 7.
 $[S']$ = Scattering matrix of 4-terminal network with self-elements S'_{11} , S'_{22} and mutual element S'_{12} .
 $[S'_{11}], [S'_{22}], [S'_{12}]$ = Moduli (magnitudes) of S'_{11} , S'_{22} , S'_{12} .
 $[S]$ = Scattering matrix of 4-terminal network consisting of $[S']$ followed by element $[\Gamma_s]$ (see Fig. 5).
 $[\bar{S}]$ = Scattering matrix of 4-terminal network consisting of $[S]$ and a transmission line of length $(-S_R)$.
 T = Designation of terminal plane identified uniquely by subscript, prime or the like.
 α = Attenuation coefficient, nepers/unit length.
 β = Phase-change coefficient, radians/unit length.

γ = A parameter of the tangent relation network of Fig. 7.

Γ = Voltage reflection coefficient.

$|\Gamma_s|$ = Value of reflection coefficient transformer (see Fig. 2).

Γ_s = See eqn. (6) and Fig. 6.

θ = Argument (or phase) of the input reflection coefficient of a 4-terminal network.

θ_0 = Angular co-ordinate in the reflection coefficient plane of centre of input locus.

$\kappa = 2\pi/\lambda_g$ = Propagation wave number.

λ_g = Guide wavelength.

ϕ = Argument (or phase) of the reflection coefficient of a termination.

$\phi_{11}, \phi_{22}, \phi_{12}$ = Arguments (or phases) of S_{11}, S_{22}, S_{12} .

(1) INTRODUCTION

In recent years, the losses inherent in plunger-type variable 'short-circuits' have become of greater interest to many microwave engineers. This tendency has had its impetus both in the search for greater accuracy in general, as well as in the development of open transmission lines, which, as a rule, do not permit the construction of lossless short-circuits.

In this connection, work has been done by G. A. Deschamps,¹ who, by means of his hyperbolic protractor, has made possible a variety of measurements involving lossy variable terminations. The term 'lossy variable termination' will be used here to designate any physical termination of constant reflection-coefficient magnitude but of variable reflection-coefficient phase. No doubt some of the material presented will overlap with his in terms of objective; however, the methods involved are quite different from his in that the less familiar non-Euclidian geometry and the hyperbolic protractor have not been used. The approach here is a straightforward analytical one. It assumes that one already knows how to measure with a perfectly lossless variable short-circuit by one of the existing techniques for obtaining the parameters of various 4-terminal representations. These techniques include, for example, Deschamps's geometric constructions² for obtaining the scattering matrix elements, and the method described by Marcuvitz³ for measuring the tangent relation parameters of a lossless structure. It is pointed out in this paper how, in a direct manner, these known techniques can be employed or modified when the short-circuit is dissipative or when any lossy variable termination is employed, and how certain measurements can be made without reference to any 4-terminal representation.

In the course of some microwave measurements the use of lossy variable terminations (such as open-ended lines, or matched lines with sliding discontinuities) becomes an undesirable necessity. This, as a rule, occurs in the case of surface-wave lines or 'microstrip' lines. On the other hand, the deliberate use of lossy variable terminations in conjunction with the measurement of low-loss quantities permits the otherwise excessively low values of voltage standing-wave ratio to be increased so that they fall into a higher range where they are measured more readily. An attenuation-coefficient measurement depending, in

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Altschuler and Prof. Oliner are at the Microwave Research Institute of the Polytechnic Institute of Brooklyn, New York.

part, on the loss in the short-circuit has already been described in the literature.⁴ In either case, it is highly desirable that the loss introduced by the termination be eliminated from the result of the measurement in a simple systematic fashion. Although the viewpoints in these two cases differ, it is readily seen that the technical approach may be the same. The applications of measurement methods involving variable lossy terminations may therefore be thought of either as correction schemes or as techniques of measuring low-loss quantities.

Here the main approach to this problem is predicated on the recognition that a lossy variable termination may be represented as a purely lossy 4-terminal network, namely, a 'reflection-coefficient transformer'* terminated in a purely lossless short-circuit. This representation permits the analysis of the measured data to proceed in a manner identical to that employed when lossless short-circuits are used. It results in a composite 4-terminal network comprised of the unknown in tandem with the 'reflection-coefficient transformer'; as will be shown, the desired parameters can be separated from this composite by using simple calculations. These methods have been applied successfully to a number of practical measurement problems. The authors have employed them to date for the following measurements:

(a) Measurements of the equivalent circuit parameters of low-loss E- and H-plane radiating slots in rectangular waveguide; a variable short-circuiting plunger was deliberately made lossy to improve accuracy.

(b) Measurements of the attenuation coefficient of short lengths of rectangular waveguide; the measurements were made feasible (for the equipment on hand) by the deliberate addition of loss to a variable short-circuiting plunger.

(c) Measurement of the equivalent circuit parameters of coaxial-line discontinuity structures; the errors introduced by the inherent loss in the coaxial-line variable short-circuiting plunger were corrected for by these techniques.

As a check on the accuracy of these methods, the equivalent circuit parameters of lossless apertures in rectangular waveguide were measured both with a good variable short-circuiting plunger and with a lossy variable termination; the results were in excellent agreement.

In addition to their use by the authors, these methods have been utilized by M. Sucher of the Microwave Research Institute for the measurement of small attenuations, and by M. G. Chernin of the Hughes Aircraft Company for the measurement of low conductance radiating slots at millimetre wavelengths.

REPRESENTATION OF LOSSY VARIABLE TERMINATIONS

The term 'lossy variable termination', as used in the paper, assumes that this structure, as regarded from a fixed terminal plane, may be represented by a lossless transmission line of variable length terminated in a fixed complex load, as shown

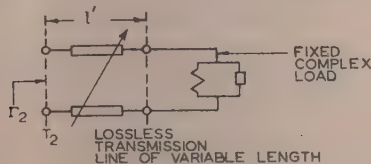


Fig. 1.—Representation of a lossy variable termination by a transmission line of variable length terminated in a fixed load.

In Fig. 1. It is evident that Γ_2 , the input reflection coefficient at the fixed plane T_2 , varies only in phase and that its magnitude $|\Gamma_2|$ is independent of l' . It is indicated, therefore, to employ a known but little used alternative representation, shown in Fig. 2,

* The reflection-coefficient transformer $|\Gamma_s|$ relates input and output reflection coefficients by $\Gamma_2 = |\Gamma_s| \Gamma_0$, in analogy to the impedance transformer relation $Z = n^2 Z_0$. Some additional details are available.⁵

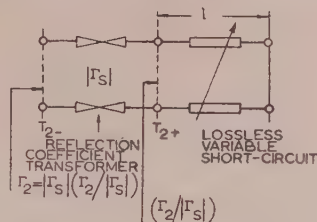


Fig. 2.—Lossy variable termination represented by a reflection-coefficient transformer terminated in a lossless variable short-circuit.

which consists of a reflection-coefficient transformer (i.e. a bilaterally matched symmetric attenuator) terminated in a lossless variable short-circuit. In view of the susceptible element in the load in Fig. 1, l' and l (in Fig. 2) are related by an additive constant. The locus of Γ_2 in both representations (as l or l' is varied) is a circle of radius $|\Gamma_s|$ with its centre at the origin of the reflection-coefficient plane.

Whereas in the course of some measurements neither $|\Gamma_s|$ nor l need be known, it is often either convenient or necessary to know one or both of these quantities. The simplest circumstances under which such measurements can be carried out are obtained if a slotted line of the same guide cross-section as that of the variable termination is available. $|\Gamma_s|$ is the reflection coefficient corresponding to r_s , the voltage standing-wave ratio* in the slotted line when it is terminated in the lossy variable termination (without any intervening discontinuities). When r_s is extremely low, one of the special techniques^{6,7} available for such measurements must be employed. The relation between $|\Gamma_s|$ and r_s is, of course,

$$|\Gamma_s| = \frac{1 - r_s}{1 + r_s} \quad (1)$$

The distance l must be related to S , the scale reading associated with the lossy termination. Let it be assumed that S is measured (to the right) from some fixed, arbitrary, but unknown, reference $S = 0$. First, the slotted line is terminated in a perfect short-circuit [see Fig. 3(a)], and D_R , the position of the probe at a

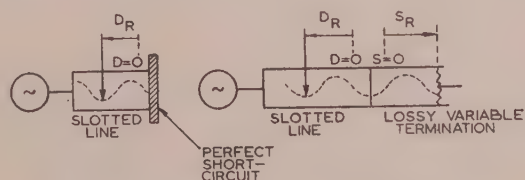


Fig. 3.—Measurement of S_R under simple conditions.

voltage minimum, is measured (to the left) from an arbitrary reference $D = 0$. The short-circuit is then replaced by the lossy variable termination [as in Fig. 3(b)] which is adjusted until a voltage minimum again occurs at D_R . When the short-circuit is in this position, the reading on the S scale is S_R . The length l , corresponding to an arbitrary setting of S , is then given by

$$l = S - S_R \quad (2)$$

When the simple calibration scheme given above cannot be employed, other means, which are described in Section 8, must be used.

Among the various physical realizations of a lossy variable termination are variable plunger-type 'short-circuits' with

* V.S.W.R. has been defined in the paper in accordance with British convention, i.e. V_{\min}/V_{\max} . This is the reciprocal of the ratio generally used in the United States.

dissipation, open-ended transmission lines which are shortened stepwise by cutting small portions from the end, and matched transmission lines along which otherwise fixed discontinuities can be moved. The open-ended transmission line has been employed in 'microstrip' line or cable measurements, for example, under the assumption that the open (radiating) end constitutes a constant load. The movable discontinuity may be employed in conjunction with a matched surface-wave line, where it might take the form of a sliding dielectric bead, or small metallic disc. In conjunction with the measurement of low-loss quantities, it may be desirable to make a variable short-circuit deliberately lossy in order to raise the measured values of v.s.w.r. into a measurable range. This can be accomplished by affixing some dissipative material firmly to the short-circuiting plunger in one of many ways.

(3) MEASUREMENT OF ARBITRARY 4-TERMINAL NETWORKS

The analyses given are perfectly general and, as such, are applicable to such more specialized 4-terminal networks as lossless or symmetric ones. In these cases, simpler procedures are obtained when the pertinent special properties are considered directly. They are treated separately in Sections 4 and 5.

(3.1) Method involving a Known Lossy Variable Termination

Let it be assumed that input-output measurements involving a lossy variable termination have been made in standard fashion (by measuring the v.s.w.r. and the position of the voltage minimum as the position of the termination is varied), by employment of the usual experimental arrangement shown in Fig. 4, and that the input data (various values of Γ_i) at T_1 have

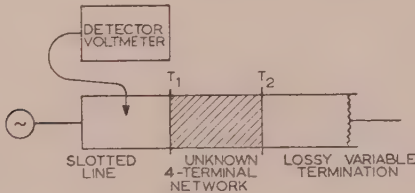


Fig. 4.—Experimental arrangement for 4-terminal network measurement.

been obtained. Then, if the unknown 4-terminal network is characterized by the scattering matrix $[S']$, one may represent the combination of the network and the lossy termination either as in Fig. 5(a) or as in Fig. 5(b). The matrix $[S]$ repre-

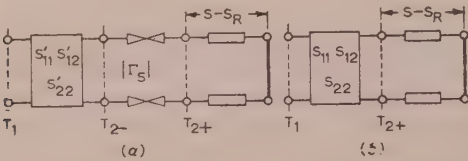


Fig. 5.—Definition of the composite 4-terminal network $[S]$.

sents a 4-terminal network which, though fictitious, is composed of the network $[S']$ and the element $|\Gamma_s|$ in tandem. The elements of the matrix $[S]$ are readily determined by an analysis of the measured data under the assumption that the termination is lossless.^{2,8} By employing the input/output relations for scattering 4-terminal networks, one can readily show that,

$$S_{11} = S_{11}, S'_{22} = S_{22}/|\Gamma_s|, S'_{12} = S_{12}/|\Gamma_s| \quad (3)$$

The measurement of a 4-terminal network with a *lossy* termination can consequently be carried out in the usual fashion and the resulting data can be analysed to obtain the network $[S]$ as if the termination had been a perfect short-circuit. The scattering parameters of interest (S'_{11} , S'_{12} and S'_{22}) are then obtained immediately from eqn. (3) provided that $|\Gamma_s|$ is known.

$|\Gamma_s|$ and S_R , both of which are constants of the termination, must be measured independently and can be obtained once and for all (at any one frequency) by a single measurement. This scheme is advantageous when a number of measurements at this frequency are contemplated with the termination in question.

(3.2) Method not requiring Knowledge of the Parameters of the Lossy Variable Termination

When only a single measurement at a given frequency is to be made, the alternative scheme based on an additional input reflection-coefficient measurement may be employed. This procedure avoids the necessity of measuring $|\Gamma_s|$ and S_R .

A completely lossless short-circuit located exactly at T_2 gives rise to a measurable value of input reflection coefficient Γ_{sc} at T_1 . With this termination, the input-output relation of the unknown 4-terminal network $[S']$ becomes

$$\Gamma_{sc} = S'_{11} - \frac{S'^2_{12}}{1 + S'_{22}} \quad (4)$$

Fig. 5 may now be redrawn as shown in Fig. 6 by associating a transmission line of length $(-S_R)$ with $|\Gamma_s|$, rather than with S .

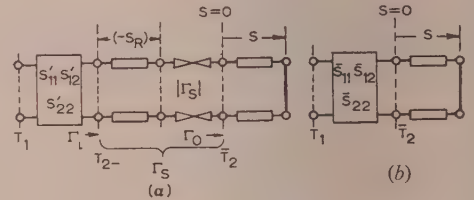


Fig. 6.—Definition of Γ_s and of 4-terminal network $[S]$.

The network $[S]$ of Fig. 6(b) includes $[S']$, $|\Gamma_s|$ and the line $(-S_R)$ in tandem and is defined with respect to T_2 , an *unknown* terminal plane. In effect, $[S]$ is the network that results when data are analysed under the assumption that the variable termination is lossless and that S , the reading on a scale with an arbitrary zero, may be taken directly to give the distance from the unknown plane T_2 to the electrical short-circuit (i.e. to the voltage null). Defining Γ_s as shown in Fig. 6(a), one may write

$$\Gamma_i = \Gamma_s \Gamma_0, \Gamma_s = e^{j2\kappa S_R} |\Gamma_s| \quad (5)$$

and eqns. (3) become

$$S'_{11} = S_{11}, S'_{22} = S_{22}/\Gamma_s, S'^2_{12} = S'^2_{12}/\Gamma_s \quad (6)$$

Eqns. (6) and (4) are solved simultaneously for the unknowns S'_{11} , S'_{22} , S'^2_{12} , and Γ_s which, in terms of the measured quantities S_{11} , S_{22} , S'^2_{12} , and Γ_{sc} , are found to be

$$S'_{11} = S_{11}, S'_{22} = 1/\left[\frac{S'^2_{12}}{S_{22}(S_{11} - \Gamma_{sc})} - 1\right], S'^2_{12} = S'_{22} S'^2_{12}/S_{22} \quad (7)$$

Γ_s can be found, if desired, from

$$\Gamma_s = \frac{S'^2_{12}}{S_{11} - \Gamma_{sc}} - S_{22} \quad (8)$$

An analysis entirely analogous to the one just given can also be carried out for the impedance representation. The expressions obtained¹⁰ are somewhat more complicated than in the scattering case and are not included here.

MEASUREMENT OF LOSSLESS 4-TERMINAL NETWORKS

Three methods of measuring lossless 4-terminal networks with a lossy variable termination are described below. The first depends on the locus in the reflection-coefficient plane and requires the measurement of the input v.s.w.r. The second is particularly suited to the measurement of discontinuities with high values of insertion v.s.w.r. and does not require that any measurements of v.s.w.r. be performed. The third is essentially a specialization of eqns. (3) to the lossless case.

One of the most convenient representations of a lossless (4-terminal) structure is that obtained from the tangent relation analysis.³ This representation, which is shown in Fig. 7, is

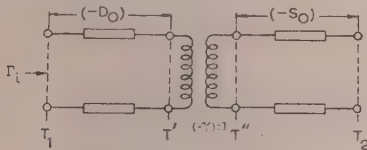


Fig. 7.—Lossless 4-terminal network represented by tangent parameters.

given directly by D_0 , S_0 and γ , the parameters derived from the analysis. γ is always negative real, with (according to the usual definition) a magnitude greater than unity. $\sqrt{-\gamma} : 1$ indicates the transformer turns ratio. The 'negative' lengths of transmission lines $-D_0$ and $-S_0$ may be alternatively thought of as $(\lambda_g/2 - D_0)$ and $(\lambda_g/2 - S_0)$.

(4.1) Method Based on the Product-Ratio Theorem

The measurement arrangement is again that shown in Fig. 4. In measurement of the input reflection coefficients Γ_i at T_1 , corresponding to various settings S of the termination [see Fig. 5(a)], the circular locus of Γ_i may be plotted. Typical loci are shown in Fig. 8. The value of $-\gamma$ is readily found from the

Γ_i -PLANE

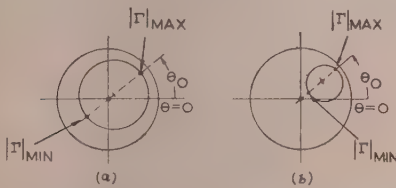


Fig. 8.—Typical loci in the input reflection coefficient plane.

well-known product-ratio theorem.* $-\gamma$ is the inverse of the insertion v.s.w.r. of the lossless 4-terminal network, while r_s is the v.s.w.r. of the variable lossy termination. Their product r_{max} (or its inverse, $1/r_{max}$) and their ratio r_{min} correspond, respectively to $|\Gamma|_{min}$ and $|\Gamma|_{max}$. One obtains

$$-\gamma = \sqrt{(r_{max}/r_{min})}; r_s = \sqrt{r_{max}r_{min}} \quad (9a)$$

as shown in Fig. 8(a), the data circle encloses the origin, and

$$-\gamma = \sqrt{(1/r_{max}r_{min})}; r_s = \sqrt{(r_{min}/r_{max})} \quad (9b)$$

the data circle does not include the origin of the reflection coefficient plane, as in Fig. 8(b). The following limiting case illustrates one of the means by which eqns. (9a) and (9b) may be distinguished from each other: When the lossy termination approaches a match, i.e. when r_s approaches unity, the data

circle degenerates to a single point (which, of course, does not include the origin). In the limit $r_{max} = r_{min} \neq 1$, while $r_s = 1$, so that eqn. (9b) is the pertinent relation for the case of the circle not enclosing the origin.

The value of D_0 is given by

$$D_0 = \frac{\lambda_g \theta_0}{4\pi} \quad (10)$$

where θ_0 is the angular co-ordinate of the centre of the data circle (see Fig. 8). To justify eqn. (10), first consider the representation in Fig. 7 when terminated in a matched load. The reflection coefficient at the input terminal plane T' of the transformer is a real positive quantity so that a voltage maximum exists at this plane. The probe in the slotted line (to the left of T_1), however, is located at a voltage minimum (some distance D_M from T_1) so that the probe is separated from T' by $\lambda_g/4$ (within an integral number of half wavelengths). One can consequently write

$$D_0 = D_M + \lambda_g/4 \quad (11)$$

It can be shown that, in the case of the lossless 4-terminal network, the centre of the data circle in the Γ_i -plane lies on the same radial line (i.e. has the same argument θ_0) as the reflection coefficient (at T_1) caused by a matched termination (at T_2). Consequently,

$$\theta_0 = \frac{4\pi D_M}{\lambda_g} - \pi \quad (12)$$

Eqns. (11) and (12) together then yield eqn. (10).

Parameter S_0 can be found only by means of an additional measurement. Let the lossless 4-terminal network be terminated by a perfect short-circuit located exactly at T_2 , and let the corresponding position of the minimum in the slotted line (to the left of T_1) be D_{sc} . By looking to the right and to the left of the plane T' one then finds that

$$-\tan \kappa (D_{sc} - D_0) = -\gamma \tan \kappa (-S_0) \quad (13)$$

Inversion of eqn. (13) gives S_0 as

$$S_0 = \frac{1}{\kappa} \arctan \left[\frac{1}{\gamma} \tan \kappa (D_0 - D_{sc}) \right] \quad (14)$$

The analysis indicated may be carried out to yield very precise results. In particular, θ_0 , $|\Gamma|_{max}$ and $|\Gamma|_{min}$ can be determined by precision averaging techniques which are already known.⁷ D_{sc} , although determined by only a single measurement, can be expected to be very accurate since only the position of a voltage null need be measured.

At the expense of accuracy, the procedure can be simplified. r_{max} , r_{min} and D_M can be determined directly by observing the input v.s.w.r. as the termination is varied. D_M is the value of D associated with r_{min} . The network parameters follow from eqns. (14), (16) and (19).

(4.2) Method Suitable for 4-Terminal Networks with Small Associated Insertion V.S.W.R.

When $-\gamma$ is almost equal to unity, i.e. when the lossless 4-terminal network presents only a small discontinuity, a fundamentally different analysis, which is much more closely related to the tangent method, can be employed advantageously. Although this method requires the separate measurement of $|\Gamma_s|$, it avoids the otherwise necessary measurement of the values of input v.s.w.r. of the lossless 4-terminal network (when the latter is terminated in the lossy termination). The derivation of this method is closely related to one already given in connection with the correction for the error introduced by an intervening lossless junction into the measured position of the voltage minimum resulting from a complex unknown load.⁹

* If a lossless 4-terminal network of insertion v.s.w.r. r_1 is terminated in a reflection coefficient of phase ϕ and associated fixed v.s.w.r. r_2 , the input v.s.w.r. takes on all values between r_1/r_2 (or r_2/r_1) and $r_1 r_2$, as ϕ is varied through all possible values.

The tangent relation, which is usually written as

$$\tan \kappa (D - D_o) = \gamma \tan \kappa (S - S_o) \quad (15)$$

is also valid in complex form, i.e. when one writes $(D + jD_i)$ instead of D and $(S + jS_i)$ instead of S . For simplicity, $D = 0$ and $S = 0$ are assumed to be located at the input and output reference planes of the structure, respectively. Under these conditions,

$$r = \tanh \kappa D_i; r_s = \tanh \kappa (-S_i) \quad (16)$$

where r is the measured input v.s.w.r. and r_s is the v.s.w.r. of the short-circuit. When $\kappa(D + jD_i - D_o)$ is abbreviated as y and $\kappa(S + jS_i - S_o)$ as x , the complex tangent relation may be written as

$$\tan y = \gamma \tan x \quad (17)$$

Let $\tan(x + y)$ now be expanded in terms of $\tan x$ and $\tan y$, and let $\tan y$ be eliminated by the use of eqn. (17); then, on defining $-\gamma \equiv 1 + \epsilon$, we obtain

$$\tan(x + y) = \frac{-\epsilon \tan x}{1 + (1 + \epsilon) \tan^2 x} \quad (18)$$

The ϵ term in the denominator may be neglected if $\epsilon \ll 1$, unless $\tan^2 x \simeq -1$. It can be shown that the restriction $(1/r_s - 1) \gg \epsilon$ is sufficient to exclude the possibility of having $\tan^2 x \simeq 1$. Since this is generally the case in practice, the only effective restriction on the applicability of this measurement is $\epsilon \ll 1$. On this assumption, the right-hand side of eqn. (18) becomes $-(\epsilon/2) \sin 2x$, with an upper limit of $\epsilon/2$, and since $\tan(x + y)$ cannot exceed the small quantity $\epsilon/2$, it may be approximated by $(x + y)$. On taking the real part of the equation so modified it is readily shown that

$$(D + S) = (D_o + S_o) - \left(\frac{\epsilon}{2\kappa} \frac{1 + r_s^2}{1 - r_s^2} \right) \sin 2\kappa(S - S_o) \quad (19)$$

where $\cosh 2\kappa S_i$ has been expressed in terms of r_s from eqn. (16) as

$$\cosh 2\kappa S_i = \frac{1 + r_s^2}{1 - r_s^2} \quad (20)$$

Eqn. (19) is the analytic expression for the $(D + S)$ versus S curve (shown in Fig. 9) of a lossless 4-terminal network charac-

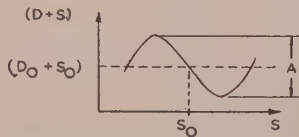


Fig. 9.— $(D + S)$ versus S curve of a lossless 4-terminal network measured with a lossy variable termination.

$$A = \left(\frac{\epsilon}{2\kappa} \frac{1 + r_s^2}{1 - r_s^2} \right)$$

terized by D_o , S_o , γ ($-\gamma \equiv 1 + \epsilon$), obtained when a lossy variable termination with v.s.w.r. r_s has been employed in the measurement. It is seen from Fig. 9 [or eqn. (19)] that S_o , $(D_o + S_o)$ and A can be abstracted from the curve at once. In terms of the amplitude A of the curve, γ is given by

$$-\gamma = 1 + 2\kappa A \frac{1 - r_s^2}{1 + r_s^2} \quad (21)$$

(4.3) Method Based on the Scattering Matrix Coefficients

The simplifications which occur when the lossless structure is to be measured by means of a scattering formalism are also worth indicating. It is known that for a lossless structure $[S']$ the scattering coefficients are restricted by

$$|S'_{11}|^2 = |S'_{22}|^2 = 1 - |S'_{12}|^2; 2\phi'_{12} = \phi'_{11} + \phi'_{22} \pm \pi \quad (22)$$

where ϕ'_{nm} is the argument of S'_{nm} . These relations, of course, do not pertain to the composite 4-terminal network $[S]$ of Fig. 5(b) since it is dissipative. On applying eqn. (3) to eqn. (22) we obtain the parameters of the unknown network $[S']$ in terms of $|S_{11}|$, ϕ_{11} and ϕ_{22} without the knowledge of $|\Gamma_s|$:

$$S'_{11} = |S_{11}|e^{j\phi_{11}}, S'_{22} = |S_{11}|e^{j\phi_{22}}, S'^2_{12} = (|S_{11}|^2 - 1)e^{j(\phi_{11} + \phi_{22})} \quad (23)$$

(5) MEASUREMENT OF SYMMETRIC 4-TERMINAL NETWORKS

When it is known that the structure of interest is capable of being represented as a symmetric 4-terminal network, the measurement can be simplified considerably. In particular, the location of the output reference plane need not be known and $|\Gamma_s|$ need not be measured.

When the condition for a symmetric representation of the scattering 4-terminal network $[S']$, namely $S'_{11} = S'_{22}$, is coupled with eqn. (6), we can immediately write

$$S'_{11} = S'_{22} = \mathcal{S}_{11} \quad (24)$$

Furthermore, on considering the ratio

$$S'^2_{12}/S'_{22} = \mathcal{S}^2_{12}/\mathcal{S}_{22} \quad (25)$$

we obtain

$$S'^2_{12} = \mathcal{S}^2_{12}\mathcal{S}_{11}/\mathcal{S}_{22} \quad (26)$$

Eqns. (24) and (26) give the desired parameters of $[S']$. Γ_s [see Fig. 6 and eqns. (5)] can also be obtained, if desired, from

$$\Gamma_s = \mathcal{S}_{22}/\mathcal{S}_{11} \quad (27)$$

upon dividing eqn. (26) by the expression for S'^2_{12} in eqns. (6). Similar relations may be found for the impedance representation.¹⁰

(6) MEASUREMENT OF ATTENUATION

An attenuator can, of course, be considered as a 4-terminal network, so that once it has been measured as such, all the values of attenuation can be abstracted in the standard fashion. If the unknown is characterized by $[S']$, we may employ the usual relations given in eqn. (28) to obtain the attenuation in decibels:

$$\left. \begin{aligned} A'_T &= -10 \log_{10} |S'_{12}|^2 &&= \text{Total attenuation} \\ A'_D &= -10 \log_{10} \frac{|S'_{12}|^2}{1 - |S'_{11}|^2} &&= \text{Reflective attenuation} \\ A'_R &= A'_T - A'_D &&= \text{Dissipative attenuation} \end{aligned} \right\} \quad (28)$$

It is to be noted that the three attenuations are always related as indicated in eqns. (28). If a lossless short-circuit is employed, $[S']$ is, of course, obtained directly from the analysis of data.

(6.1) Method Suitable for Arbitrary Attenuators

When the termination involved is lossy, $[S']$ can be obtained from eqns. (6) or (7). In making this measurement the attenuator is placed between the slotted line and the lossy termination so that the power is incident on it in the same direction as when it is in actual use. Since only the magnitudes of the scattering coefficients are required in eqns. (28), it is sufficient to measure only $|S_{11}|$ and $|S_{12}|$, as defined in connection with eqns. (6). This makes the additional experimental determination of the output reference plane unnecessary. It is pertinent to stress again that loss in the termination may well increase the accuracy of the measurement of small attenuations (and other low-loss quantities), so that its deliberate addition to a short-circuit may be indicated. Some considerations of this point can be found elsewhere.¹⁰

(6.2) Method Suitable for Symmetric Attenuators

When the attenuator can be assumed to be symmetric, eqns. (28) may be rewritten by the use of eqns. (24) and (26) to yield the desired result:

$$A'_T = -10 \log_{10} \frac{|\bar{S}_{12}|^2 |\bar{S}_{11}|}{|\bar{S}_{22}|},$$

$$A'_D = -10 \log_{10} \frac{|\bar{S}_{12}|^2 |\bar{S}_{11}|}{|\bar{S}_{22}|(1 - |\bar{S}_{11}|^2)} \quad (29)$$

Here $|\bar{S}_{11}|$, $|\bar{S}_{22}|$ and $|\bar{S}_{12}|$ are obtained directly from the data and $|\Gamma_s|$ need not be known.

(6.3) Alternative Method Suitable for Arbitrary Attenuators

When the interest lies only in the attenuation of a structure rather than in the parameters of $[S']$ *per se*, we can modify the procedure of Cullen¹¹ or Beatty.¹² The attenuator is measured as before, except that it is placed between the slotted line and the lossy variable termination in reverse so that the power is incident on it in the direction *opposite* to that in actual use. If the parameters of the composite structure consisting of the reversed attenuator and the reflection coefficient transformer $[\Gamma_s]$ are given by $[\sigma]$, and those of the reversed attenuator alone by $[\sigma']$, then $[\sigma]$ and $[\sigma']$ are related in the manner of eqns. (3). On account of the reversal we have

$$\sigma_{11} = S'_{22}; \sigma'_{12} = S'_{12}; \sigma'_{22} = S'_{11} \quad (30)$$

It has been shown^{11,12} that A'_D , the dissipative attenuation of a structure $[S']$, is given by

$$A'_D = -10 \log_{10} R \quad (31)$$

where R is the radius of the input locus when the reversed structure is measured with a *perfect* short-circuit. On comparing with eqns. (28) and applying eqns. (30) we find that

$$R = \frac{|\sigma'_{12}|^2}{1 - |\sigma'_{11}|^2} = \frac{|\sigma'_{12}|^2}{1 - |\sigma'_{22}|^2} \quad (32)$$

where $[S']$ refers to the unreversed structure and $[\sigma']$ to the structure as it is actually being measured. It follows that, when the reversed structure is being measured with a lossy variable termination, the composite $[\sigma]$ is related to ρ , the radius of the corresponding input locus, in the same manner as $[\sigma']$ is related to R in eqns. (32), so that

$$\rho = \frac{|\sigma_{12}|^2}{1 - |\sigma_{22}|^2} \text{ or } |\sigma_{12}|^2 = \rho(1 - |\sigma_{22}|^2) \quad (33)$$

Eqns. (28) may now be expressed in terms of $[\sigma]$ from eqns. (30) and (3):

$$A'_T = -10 \log_{10} \frac{|\sigma_{12}|^2}{|\Gamma_s|}, A'_D = -10 \log_{10} \frac{|\Gamma_s| |\sigma_{12}|^2}{|\Gamma_s|^2 - |\sigma_{22}|^2} \quad (34)$$

When, in addition, $|\sigma_{12}|^2$ is eliminated from eqns. (34) by means of eqn. (33), we have the desired result

$$A'_T = -10 \log_{10} \frac{\rho(1 - |\sigma_{22}|^2)}{|\Gamma_s|},$$

$$A'_D = -10 \log_{10} \frac{\rho |\Gamma_s| (1 - |\sigma_{22}|^2)}{|\Gamma_s|^2 - |\sigma_{22}|^2} \quad (35)$$

Although $|\Gamma_s|$ may be found by means of a separate, direct measurement, it can also be found in the manner of eqn. (8) by making an additional single-point measurement and computing

$$|\Gamma_s| = \left| \frac{\sigma_{12}^2}{\sigma_{11} - \gamma_{sc}} - \sigma_{22} \right| \quad (36)$$

if the attenuator is a symmetric structure, it may be obtained from eqn. (27)] from

$$|\Gamma_s| = |\sigma_{22}|/|\sigma_{11}| \quad (37)$$

The quantity γ_{sc} in eqn. (36) is the input reflection coefficient measured when the reversed attenuator is terminated in a perfect short-circuit.

Eqns. (35), which are based on the measurement with the attenuator reversed, are relatively insensitive to errors in $|\sigma_{22}|$. In this connection [provided that eqn. (36) or (37) is not employed], both $|\Gamma_s|$ and ρ can be found as the radii of circles in a very precise manner,⁷ so that when $|\sigma_{22}|$ is relatively small, as it normally is in this type of measurement, eqns. (35) usually yield reliable results.

(7) MEASUREMENT OF ATTENUATION COEFFICIENT

In the measurement of the attenuation coefficient of waveguide, the power dissipated in the sample guide is often so small that the input v.s.w.r. of the short-circuited sample guide is excessively low and lies outside the useful (stable) region of the detecting device employed. Under these circumstances, a lossy variable termination may be employed to increase the observed v.s.w.r. to lie in a stable region of the equipment. It may be shown that, assuming no problem of instability, a measurement may be made more sensitive when the point of measurement is shifted to the higher portion of the appropriate v.s.w.r. meter scales by the addition of loss to the short-circuit.¹⁰

Once it is assumed that the variable termination is lossy, we can categorize the measurements of attenuation coefficient according to the specific locations of the discontinuities involved, if any, and treat each such case separately. This is done below. A basic arrangement, such as the one shown in Fig. 4, consisting of a sample waveguide located between a slotted line and a lossy variable termination, is assumed. The input reflection coefficients corresponding to a number of different termination settings are measured, and the measurement is then repeated with the sample line removed and with the lossy variable termination connected directly to the slotted line. Attenuation coefficients are then abstracted from the data so obtained in the manner most suited to the case at hand.

(7.1) Attenuation-Coefficient Measurement involving no Discontinuities

When it is assumed that no discontinuities are present between the probe and the termination, the procedure is very direct. Although this measurement can be performed by the use of single-point data only, increased accuracy can be expected when complete data circles are employed. When no discontinuities are present, these circles are centred about the origin, so that, in connection with the measurements both with and without the sample in place, an average of the values of v.s.w.r. observed (as the termination takes on various values) is of interest. If one defines $|\Gamma_s|$ and $|\underline{\Gamma}_s|$ as the corresponding magnitudes of the reflection coefficients determined without and with the sample line, respectively, we have simply

$$\alpha = \frac{1}{2d} \log_e [|\Gamma_s|/|\underline{\Gamma}_s|] \text{ nepers/unit length} \quad (38)$$

This relation follows directly when one considers that $|\underline{\Gamma}_s|$ is the magnitude of the reflection coefficient obtained when $|\Gamma_s|$ is observed through a transmission line of length d and of attenuation coefficient α .

(7.2) Attenuation-Coefficient Measurement involving a Discontinuity associated with the Lossy Variable Termination

The method described here applies when the measurements, both with and without the sample in place, involve an arbitrary discontinuity associated with the lossy variable termination. For example, this is the case when a slotted line of the same guide

cross-section as the sample guide itself is employed in conjunction with a termination of different cross-section. The same adaptor (discontinuity) would be joined to the variable termination during both measurements. Both sets of data obtained (i.e. with and without the sample) are plotted as circles in the reflection-coefficient plane, and the maximum magnitude of reflection coefficient on each circle is determined. When these are defined, respectively, as $|\Gamma|_{\max}$ and $|\Gamma|_{\max}$ with and without the sample waveguide in place, the attenuation coefficient is given by

$$\alpha = \frac{1}{2l} \log_e \frac{|\Gamma|_{\max}}{|\Gamma|_{\max}} \text{ nepers/unit length} \quad (39)$$

As in the previous case, this relation follows once it is recognized that the input locus obtained without the sample waveguide is modified only by a multiplicative factor of

$$e^{-2\alpha d} e^{-2j\beta d}$$

to yield the locus obtained with the sample waveguide in place.

(7.3) Attenuation-Coefficient Measurement involving a Discontinuity associated with the Slotted Line

When an arbitrary discontinuity is associated with the slotted line (in the same sense as a discontinuity was associated with the lossy variable termination above) the method described here applies. If the arbitrary discontinuity is thought of as a junction $[S']$, the measurement without the sample line will result in the 4-terminal network $[S]$ when analysed as if the termination had been lossless (see Fig. 5). $[S]$ and $[S']$ are related by eqns. (3). Likewise, when the sample line is inserted, the 4-terminal network computed under the lossless termination assumption is taken as $[\underline{S}]$, where the lossy termination $|\Gamma_s|$ rather than the 4-terminal network $[S']$ is to be regarded as having been modified by the sample waveguide dissipation. We can consequently rewrite eqns. (3) as

$$|S'_{11}| = |\underline{S}_{11}|, |S'_{22}| = |\underline{S}_{22}| / (|\Gamma_s| e^{-2\alpha d}), \\ |S'_{12}|^2 = |\underline{S}_{12}|^2 / (|\Gamma_s| e^{-2\alpha d}) \quad (40)$$

in connection with this second measurement. Eqns. (3) and (40) at once yield the relation

$$e^{2\alpha d} = |S_{12}|^2 / |\underline{S}_{12}|^2 = |S_{22}| / |\underline{S}_{22}| \quad (41)$$

when the appropriate ratios are taken. The expression giving the attenuation coefficient in terms of the available scattering coefficients is therefore

$$\alpha = \frac{1}{2d} \log_e \frac{|S_{12}|^2}{|\underline{S}_{12}|^2} = \frac{1}{2d} \log_e \frac{|S_{22}|}{|\underline{S}_{22}|} \text{ nepers/unit length} \quad (42)$$

(7.4) Attenuation-Coefficient Measurement involving Lossless Discontinuities associated with both Slotted Line and Termination

The case treated here is somewhat more complicated than those already considered and finds a practical application in the measurement of the attenuation coefficient of cables which are joined to both the slotted line and the lossy termination by connectors (lossless discontinuities).

Here, it is of interest to examine more carefully what, exactly, constitutes a lossless discontinuity. Rigorously, even when the cross-sections of a lossy and a lossless guide are geometrically identical, a dissipative discontinuity exists at their junction regardless of whether or not a geometrical obstacle is present. In almost all practical cases such discontinuities can either be neglected or, when, in addition, a geometrical obstacle is present, can be considered to be lossless. In extreme cases, however, it may be well to consider whether or not the loss in such a dis-

continuity can affect the accuracy of the attenuation-coefficient measurement which follows.

The measurement and analysis associated with this procedure is explained with reference to Fig. 10. Data are taken with and

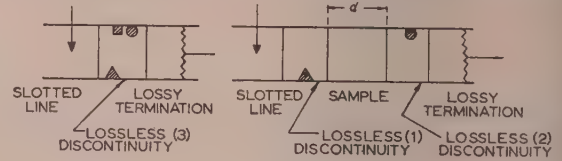


Fig. 10.—Measurement arrangement with and without sample in place.

without the sample in place, and the two input loci are plotted. As the sample is removed, discontinuities 1 and 2 may be joined to contribute to discontinuity 3; this is permissible, but not necessary, since the measurement without the sample is carried out only for the purpose of determining $|\Gamma_s|$. The relationship, if any, which discontinuity 3 bears to 1 and 2 is consequently of no interest, and $|\Gamma_s|$ may be found from the plotted locus as pointed out in connection with eqn. (9).

The locus obtained with the sample in place is first analysed by means of eqns. (3). Discontinuity 1, the sample and discontinuity 2, taken together, form the unknown 4-terminal network $[S']$ in the sense of Fig. 5. Once $[S']$ has been obtained from $[S]$ and $|\Gamma_s|$, the set of input reflection coefficients, which would be observed if the network $[S']$ were measured with a truly lossless variable short-circuit, is found by computing

$$\Gamma_i = \frac{(S'_{11} S'_{22} - S'^2_{12}) \Gamma_0 - S'_{11}}{S'_{22} \Gamma_0 - 1}; \Gamma_0 = e^{j\phi} \quad (43)$$

for various values of ϕ . It is recognized that the locus of Γ_i is precisely the same as that which would result if data had been taken with a lossless variable short-circuit and without discontinuity 2 present. This computed locus may therefore be viewed as the locus resulting from the measurement of a lossless 4-terminal network (discontinuity 1) using a lossy variable-termination composed of the sample line and the lossless short-circuit. The computed data are plotted as in Fig. 8 and analysed for r_s according to eqns. (9). From r_s we have

$$|\Gamma_s| = \frac{1 - r_s}{1 + r_s} \quad (44a)$$

But $|\Gamma_s|$ is also clearly

$$|\Gamma_s| = e^{-2\alpha d} \quad (44b)$$

where α is the attenuation coefficient of the sample. By combining eqns. (9) and (44) we can express α directly in terms of the reflection-coefficient magnitudes found from the computed locus as

$$\alpha = \frac{1}{2d} \log_e \frac{\sqrt{[(1 + |\Gamma|_{\max})(1 \pm |\Gamma|_{\min})] + \sqrt{[(1 - |\Gamma|_{\max})(1 \mp |\Gamma|_{\min})]}}}{\sqrt{[(1 + |\Gamma|_{\max})(1 \pm |\Gamma|_{\min})] - \sqrt{[(1 - |\Gamma|_{\max})(1 \mp |\Gamma|_{\min})]}}} \quad (45)$$

The upper sign is taken if the locus encloses the origin of the chart, while the lower sign pertains if it does not.

(8) CALIBRATION OF THE LOSSY VARIABLE TERMINATION

In the measurements already described, the properties of the lossy variable termination must sometimes be known in order to find the quantities of interest. In particular, means of finding both S_r and $|\Gamma_s|$ or r_s may have to be available. S_r has already been defined in connection with eqn. (2) and Figs. 2 and 3, and its experimental determination, when the slotted line can be

terminated directly in the lossy variable termination without any intervening discontinuities, has been described there.

(8.1) Measurement of the Magnitude of Γ_s

When the connection between the lossy variable termination and the slotted line contains lossless discontinuities, r_s (and from r_s , $|\Gamma_s|$) is determined readily by the product-ratio theorem in the manner given in connection with eqns. (9). This scheme is both relatively quick and capable of yielding precise results.

If the intervening discontinuity is dissipative, but is known to be symmetric with respect to some reference plane, $|\Gamma_s|$ can be found from eqn. (27). This involves the measurement of S_{11} and S_{22} , i.e. of the magnitudes of two of the elements of the composite scattering 4-terminal network $[S]$, as described in connection with Fig. 6. It should be noted that in this, as well as in the lossless case above, it is not necessary that the termination and the discontinuity be physically detachable from each other.

When the discontinuity between the slotted line and the lossy termination is dissipative and asymmetric, $|\Gamma_s|$ can be found only if one of a number of certain measurements is performed in addition to the measurement of $[S]$. One of these is described in connection with eqn. (8), which holds when the lossy variable termination is replaced by a perfect short-circuit and Γ_{sc} is measured. The resulting value of Γ_{sc} , in conjunction with the measured parameters of 4-terminal network $[S]$, then yields $|\Gamma_s|$ from eqn. (8). If the termination cannot be detached, however, $|\Gamma_s|$ must be measured directly. This can be done

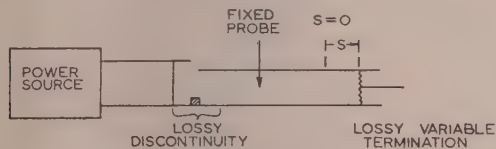


Fig. 11.—Experimental arrangement for the determination of $|\Gamma_s|$ and S_R by means of a fixed probe.

by placing a fixed probe in the transmission line between the discontinuity and the lossy variable termination, as shown in Fig. 11, and by determining the v.s.w.r. at the fixed probe as the termination is varied.

(8.2) Measurement of S_R

In general S_R can be found if either a probe or a fixed short-circuit can be located at some plane T_2 , which then serves as the put reference plane of the variable termination. If the discontinuity associated with the termination is dissipative, a probe must be used. If the discontinuity is lossless, however, either a probe or a fixed short-circuit can be employed. In view of the thickness effects of probes, a fixed short-circuit is generally preferable.

When a probe is employed to find S_R , Fig. 11 again applies. The termination is varied until a voltage minimum appears at the probe (i.e. at T_2) and the value indicated on the S -scale is noted. This value of S equals S_R .

When a variable termination is used in connection with an intervening lossless 4-terminal network (represented as in Fig. 7), S_0 and γ are first found from eqns. (9) and (10). These values in conjunction with D_{SC} , which is defined in connection with eqn. (13), are employed in eqn. (14) to yield S_0 . The necessary measurements are described in connection with the equations mentioned. Once S_0 is known, one finds S_R from

$$S_R = S_{min} - S_0 \quad (46)$$

where S_{min} is defined as the value indicated on the S -scale of the variable termination when it is in such a position that $|\Gamma_i|$ is a minimum (see Fig. 8). $|\Gamma_i|$ will be recalled as the magnitude of the input reflection coefficient of the 4-terminal network, i.e. of the intervening discontinuity. Eqn. (46) is now explained with reference to Fig. 12. As shown, S_R is the S -scale reading

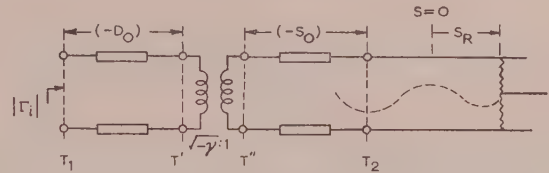


Fig. 12.—Lossless 4-terminal network terminated in a lossy variable termination.

when a voltage minimum occurs at T_2 . It can be shown that when $|\Gamma_i|$ is a minimum a voltage minimum occurs at T'' . Since these two locations of voltage minimum are separated by a transmission line of length $(-S_0)$, eqn. (46) follows.

(9) ACKNOWLEDGMENT

The work described in the paper was sponsored by the Air Force Cambridge Research Center, Air Research and Development Command, U.S. Air Force.

(10) REFERENCES

- (1) DESCHAMPS, G. A.: 'A Hyperbolic Protractor for Microwave Impedance Measurements and Other Purposes' (Federal Telecommunication Laboratories, Nutley, New Jersey, 1953).
- (2) DESCHAMPS, G. A.: 'Determination of Reflection Coefficients and Insertion Loss of a Waveguide Junction', *Journal of Applied Physics*, 1953, **24**, p. 1046.
- (3) MARCUVITZ, N.: 'On the Representation and Measurement of Waveguide Discontinuities', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 728.
- (4) VOGELMAN, J. H.: 'Precision Measurement of Waveguide Attenuation', *Electronics*, December, 1953, **26**, p. 196.
- (5) ALTSCHULER, H. M.: 'A Method of Measuring Dissipative Four-Poles Based on a Modified Wheeler Network', *Transactions of the Institute of Radio Engineers*, January, 1955, **MTT-3**, p. 30.
- (6) MONTGOMERY, C. G.: 'Technique of Microwave Measurements', Radiation Laboratory Series, **11** (McGraw-Hill, New York, 1947), pp. 505-507.
- (7) ALTSCHULER, H. M., and OLINER, A. A.: 'A Shunt Technique for Microwave Measurements', *Transactions of the Institute of Radio Engineers*, July, 1955, **MTT-3**, p. 24.
- (8) DUKES, M. C.: 'Waveguides and Waveguide Junctions', *Wireless Engineer*, 1955, **32**, p. 65.
- (9) OLINER, A. A.: 'The Calibration of the Slotted Section for Precision Microwave Measurements', *Review of Scientific Instruments*, 1954, **25**, p. 13.
- (10) ALTSCHULER, H. M., and OLINER, A. A.: 'Microwave Measurements with a Lossy Variable Short Circuit' (Microwave Research Institute, Polytechnic Institute of Brooklyn, Report R-399-54).
- (11) CULLEN, A. L.: 'Measurement of Microwave Transmission Efficiency', *Wireless Engineer*, 1949, **26**, p. 255.
- (12) BEATTY, R. W.: 'Determination of Attenuation from Impedance Measurements', *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 895.

SOME PROPERTIES OF THE TRANSFER FUNCTION OF UNBALANCED RC NETWORKS

By I. CEDERBAUM.

(The paper was first received 30th December, 1955, and in revised form 8th March, 1956. It was published as an INSTITUTION MONOGRAPH in June, 1956.)

SUMMARY

A method is presented for deducing some RC network theorems, chiefly concerning the transfer function of an unbalanced structure from the properties of pure resistance networks. This is possible mainly since:

(a) Zero of the transfer function (voltage ratio) of a 3-terminal R network has a clear topological interpretation, namely in the case of a zero transfer the network disintegrates in two parts (at least), which may have only the reference node in common, and the input and output terminals belong to different parts.

(b) The expansion of the RC network transfer function about the values of all its susceptances is a quotient of two polynomials in all these susceptances. The coefficients in this expansion are composed exclusively of cofactors of purely real matrices. These matrices correspond to the original network with some capacitors short-circuited and others open-circuited.

(c) Vanishing of some coefficients in the numerator of the expansion of the transfer function of a 3-terminal RC network implies a zero of the transfer function for some of the R networks derived in this way. This fact according to (a) imposes definite conditions upon the topology of the original RC network.

These limitations show that the only possible form of a degeneration of the transfer function in its non-reduced form is through the disappearance of some left or right consecutive terms of the numerator. For a generalized ladder realization of such degenerated transfer function the discussion offers the limits of the maximum gain.

(1) INTRODUCTION

It has been pointed out elsewhere¹ that the solution of a problem in network synthesis is not sufficiently tied to a practical network. The designer starting with his work has only a very general idea about the geometry of the network and the amount and interconnection of elements which will eventually be needed for its realization.

The best way to overcome this difficulty seems to lie in gaining more topological knowledge on the network. Any new information concerning the relation between the geometry of the network and the form of the response function should be regarded as a step forward.

An attempt is made in the paper to study the influence of the location of some element or group of elements on the network response. To this end, the response function is considered as a function of those chosen elements. This approach is not new; it is basically common to many branches of electrical engineering and has been used by Carter² and Feldtkeller³ in analysing the driving-point impedances of electrical networks. As applied to RC networks—and the same is true for RLC networks without mutual inductance—it offers the striking advantage that after picking out all the capacitances as variable parameters we are left with a structure composed of pure resistances only. The form of the response function remains unchanged if the capacitors are replaced by resistors (or other network elements) and thus the natural consequence of this method of approach would be to

start with the discussion of purely resistive or R networks. Alternatively, this method might be looked on as deducing the properties of an RC network from its behaviour in the neighbourhood of some real positive value of the complex frequency for which each capacitor may be replaced by an appropriate resistor. This point of view was stressed by Reza⁴ and Talbot.⁵

The method seems to be particularly suitable for dealing with the problem of the limits of the possible gain. Fialkov and Gerst⁶ have given these limits for the transfer function of 3- and 4-terminal RC networks when the number of elements for disposal is unlimited. But the important question of how this gain limitation depends on the number of elements remains unsolved. In what follows, some of the degenerate forms of the transfer function of 3-terminal networks with the minimum number of elements are discussed.

(2) PRELIMINARY REMARKS

Consider an RC network as consisting of a basic R structure with the capacitance components inserted in some places. In the sequel it will be shown that all cofactors of the admittance matrix of an RC network, and by means of them all RC network functions, may be expressed by the cofactors of purely real matrices which correspond to the original network with some capacitors short-circuited and others open-circuited. Such a state will be called an 'aspect' of the network.

The process of deriving the various aspects from an RC network may sometimes lead to R networks which are composed of two or more sub-networks. An example of such a situation is shown in Fig. 1; the original RC network is shown at (a) and two of its

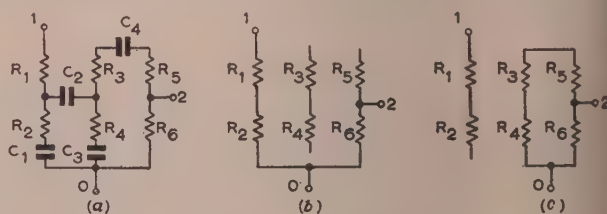


Fig. 1.—An RC network and two of its aspects.

possible aspects are shown at (b) and (c). Fig. 1(b) corresponds to short-circuiting the capacitor C_1 and open-circuiting all other capacitors, whereas Fig. 1(c) corresponds to short-circuiting C_3 and C_4 and open-circuiting C_1 and C_2 . It can be seen from these diagrams that as the result of open-circuiting or short-circuiting the capacitors of a network some of its parts may appear fully isolated from the reference node, or there may appear some parts which are connected with the rest only at this one point. In the following discussion the term 'sub-network' is applied for simplicity to each separate part which is either fully isolated or connected with the rest of the network only at the reference node. Evidently Figs. 1(b) and 1(c) correspond to networks with three and two sub-networks respectively, 0 denoting the reference node.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Cederbaum is at the Imperial College of Science and Technology, University of London (on leave of absence from the Scientific Department, Ministry of Defence, Israel).

In order to write down the conductance matrix of such a general R network with reference to some node 0, the nodes belonging to each sub-network may be denoted by consecutive indices. With this convention the conductance matrix G of the network turns out to be composed of square blocks $G_\alpha, G_\beta, \dots, G_\mu$ located on the main diagonal, each block corresponding to a different sub-network, thus

$$G = \begin{vmatrix} G_\alpha & 0 & \dots & 0 \\ 0 & G_\beta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & G_\mu \end{vmatrix} \quad (1)$$

Since there may be a connection between various sub-networks through the reference node only, all other elements of the matrix G are zeros. All the sub-matrices which correspond to the sub-networks isolated from the reference node are singular matrices, since the sums of all their elements in each row are zeros. The others are non-singular.

With regard to the transfer of energy, the most important networks are those which are connected and cannot be subdivided into two parts connected at the reference node only. Such networks are non-separable at the chosen reference node 0 or, expressed shortly, n.s.0.

In each n.s.0 3-terminal network, if we start from an arbitrary node and advance continuously through the elements of the network, we may reach any other node in general on many paths. Each path passes through some nodes and represents a chain of series-connected elements of the network. We exclude those paths which contain closed loops and also those which pass through the reference node. Call any other a transfer path. Two transfer paths are considered different if they differ in at least one element (although they may pass through the same nodes). In Fig. 2 the chains *abeikm* and *abfikm* represent two

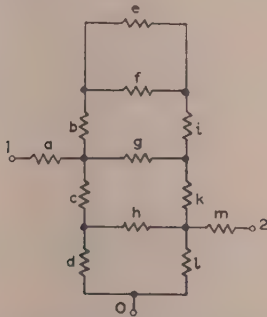


Fig. 2.—Example of an 'n.s.0' 3-terminal network.

The nodes 1 and 2 may be joined by various transfer paths such as *abeikm* and *abfikm*. The chains *abfigchm*, *agifbchm* or *acdilm* are not transfer paths.

different transfer paths between the nodes 1 and 2, whereas *abfigchm*, *agifbchm* and *acdilm* do not represent transfer paths in the sense explained above. Evidently a network is n.s.0 if, and only if, each pair of its nodes may be joined by means of at least one transfer path composed of its elements.

For connected R networks, the energy considerations lead immediately to the conclusion that the determinant of the conductance matrix and all its principal minors are positive numbers.

Denoting by T the transfer function (voltage ratio) of the terminal R network (3TRN) we have as the possible range for in the case of connected networks,⁵ $0 \leq T \leq 1$. Since the

transfer function from the node i to the node k may be expressed by the cofactors of the conductance matrix as

$$T = \frac{D_{ik}}{D_{ii}} \quad (2)$$

the following theorems are true in this case:

Theorem (1).—All cofactors D_{ik} ($i, k = 1, 2, \dots, n$) of the conductance matrix are non-negative.

Theorem (2).—The cofactor D_{ii} corresponding to the main diagonal term of the conductance matrix is not smaller than the cofactor D_{ik} corresponding to any other term in the same row (or column).

These theorems remain true for a general R network which may be not connected, since the only difference for such a network is that D_{ii} may be zero, and in that case all D_{ik} for $k = 1, 2, \dots, n$ are also equal to zero.

For the transfer function of a 3TRN the following cases may be distinguished:

(a) If the input and output nodes belong to the same sub-network, $0 \leq T \leq 1$. If this sub-network is isolated from the reference node, $T = 1$ always, since the potential of the whole of this network is identical with that of the input node. For an n.s.0 network the equality $T = 0$ cannot occur (save in the trivial case when the output terminal coincides with the reference node), since in such a network there must exist a transfer path from the input to the output terminal.

(b) If the input and output nodes belong to two different sub-networks, $T = 0$, or T is indefinite, according to whether the sub-network containing the output node is connected with the reference node or is floating.

(3) TRANSFER FUNCTION AS A FUNCTION OF ONE NETWORK CONDUCTANCE

Consider the case of a connected R network and let 1 be the input, 2 the output and 0 the reference node. This notation will be adopted throughout the paper.

Since the cofactors D_{ik} are polynomials linear in each conductance of the network, the transfer function is a linear rational function of each conductance; thus

$$T(G_1) = \frac{D_{12}(G_1)}{D_{11}(G_1)} = \frac{a + cG_1}{b + dG_1} \quad (3)$$

The coefficients a and b may be recognized as the values of the corresponding determinants when $G_1 = 0$. Similarly c and d are the values of those determinants with the element G_1 short-circuited, i.e. for the network obtained by identifying the terminals of G_1 (see Section 10.1). Taking this into account we may write

$$T(G_1) = \frac{D_{12}(o) + G_1 D_{12}(s)}{D_{11}(o) + G_1 D_{11}(s)} \quad (4)$$

where o and s indicate the two aspects of the network, with the element G_1 open-circuited and short-circuited respectively.

As a linear rational function of G_1 , continuous for $G_1 \geq 0$, the transfer function, if not constant, is strictly monotonic. As such, $T(G_1)$ may reach its extreme values only at the extreme values of G_1 ; hence

Theorem (3).—The transfer function $T(G_1)$ of a 3TRN, if not constant, may equal zero or unity only for $G_1 = 0$ or $G_1 = \infty$.

The geometry of the networks corresponding to Theorem 3 is defined thus:

Theorem (4).—If $T(G_1)$ is zero for $G_1 = 0$ or $G_1 \rightarrow \infty$ (but is not identically zero), the network may be subdivided into two parts. The input node belongs to one part and the output node to the other. In the case $T(0) = 0$ both parts are connected by means of one conductance G_1 only, and the part containing the output node must

be connected with the reference node. In the case $T(s) = 0$ [$T(s)$ being a shorthand notation for $\lim_{G_1 \rightarrow \infty} T(G_1)$] both parts have only one node in common save the possibly common reference node, and the specified conductance G_1 connects this node with the reference node. Figs. 3(a) and 3(b) illustrate the two cases $T(0) = 0$ and $T(s) = 0$ respectively.

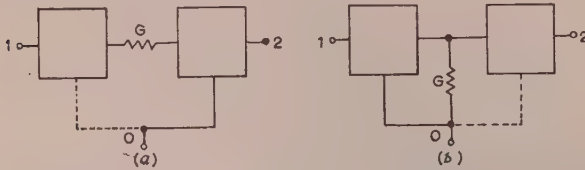


Fig. 3.—Two cases of vanishing of the transfer function $T(G)$.
(a) $T(0) = 0$; (b) $T(s) = 0$.

This may be proved in the following way. The condition $T = 0$ implies that in this case the nodes 1 and 2 belong to two different sub-networks and the sub-network containing the output node must be connected with the reference node. If a connected network can be decomposed into two such sub-networks by opening the connection between two nodes ($G_1 = 0$), the original network consists of these two parts connected by G_1 as stated. If this decomposition is the result of short-circuiting two nodes ($G_1 \rightarrow \infty$), one node of the original network must merge into the reference node as the result of G_1 tending to infinity. This is the only node common to both parts save the possible common reference node. The part containing the output node may be connected in this case with the reference node only through G_1 , since after G_1 tends to infinity it remains connected to this point.

Consider now, instead of T , the complementary transfer function \bar{T} of the same network with the roles of the input and reference nodes interchanged, the output node being unchanged. This was used by Fialkow and Gerst.⁶ Evidently $\bar{T} = 1 - T$, and if Theorem (4) be applied to the case $\bar{T} = 0$ (and $T = 1$) the further proposition can be stated thus:

Theorem (5).—If $T(G_1) = 1$ for $G_1 = 0$ or $G_1 \rightarrow \infty$ (but is not identically unity) the network can be subdivided into two parts, as in Theorem (4), with the roles of the input and reference nodes interchanged.

(4) TRANSFER FUNCTION AS A FUNCTION OF n CHOSEN NETWORK CONDUCTANCES

If the transfer function of a connected 3TRN is considered as a function of some n conductances G_1, G_2, \dots, G_n , it is a quotient of two polynomials in all these conductances. When $n = 2$, e.g. if the chosen conductances are G_1 and G_2 , we may initially express $T(G_1, G_2)$ as a function of G_2 and afterwards, proceeding in the same manner, express all the coefficients as functions of G_1 . Thus we have

$$\begin{aligned} T(G_1, G_2) &= \frac{D_{12}(G_1, G_2)}{D_{11}(G_1, G_2)} = \frac{D_{12}(G_1, o) + G_2 D_{12}(G_1, s)}{D_{11}(G_1, o) + G_2 D_{11}(G_1, s)} \\ &= \frac{D_{12}(o, o) + G_1 D_{12}(s, o) + G_2 D_{12}(o, s) + G_1 G_2 D_{12}(s, s)}{D_{11}(o, o) + G_1 D_{11}(s, o) + G_2 D_{11}(o, s) + G_1 G_2 D_{11}(s, s)} \end{aligned} \quad (5)$$

where o or s standing instead of G_1 or G_2 indicates the open-circuiting or short-circuiting of the corresponding conductance respectively.

The numerator and denominator of this expression are bilinear functions in both variables G_1 and G_2 . All terms with G_1 or G_2 have in the original cofactor the corresponding conductance replaced by s . The terms without G_1 or G_2 have o at the same place.

For the general case of n conductances it is convenient to introduce some shorthand notation. Let the symbol

$D_{gh}^k(G_1, G_2, \dots, G_n)$ represent the sum of all $\binom{n}{k}$ possible products of k out of n conductances G_1, G_2, \dots, G_n , each product multiplied by the cofactor D_{gh} taken for these k elements short-circuited and the remaining $(n - k)$ elements open-circuited.

Thus, e.g.,

$$D_{12}^1(G_1, G_2, \dots, G_n) = G_1 D_{12}(s, o, \dots, o) + G_2 D_{12}(o, s, \dots, o) + \dots + G_n D_{12}(o, o, \dots, s) \quad (6)$$

The special cases when one conductance connects a pair of the three terminals is dealt with in Section 10.1.

For $k = 0$ we define

$$D_{12}^0(G_1, G_2, \dots, G_n) = D_{12}(o, o, \dots, o) \quad (7)$$

With these conventions eqn. (5) can be rewritten

$$T(G_1, G_2) = \frac{\sum_{k=0}^2 D_{12}^k(G_1, G_2)}{\sum_{k=0}^2 D_{11}^k(G_1, G_2)} \quad (8)$$

Applying now the same method of consecutive expansions to the general case we readily get

$$T(G_1, G_2, \dots, G_n) = \frac{\sum_{k=0}^n D_{12}^k(G_1, G_2, \dots, G_n)}{\sum_{k=0}^n D_{11}^k(G_1, G_2, \dots, G_n)} \quad (9)$$

A reasoning similar to that used for proving Theorem (3) when applied to the case of n variable conductances leads to

Theorem (6).—The transfer function $T(G_1, G_2, \dots, G_n)$ of the 3TRN, if not constant, can reach the value 0, or 1, only if one, at least, of the conductances G_1, G_2, \dots, G_n is 0 or $\rightarrow \infty$.

Consider now some n conductances of a 3TRN: G_1, G_2, \dots, G_n . Suppose that a number k exists, $0 \leq k \leq n$, such that the transfer function of the 3TRN equals zero for each combination of k of those conductances short-circuited and the remaining $n - k$ open-circuited, but is not identically zero. Evidently in this case also $D_{12}^k(G_1, G_2, \dots, G_n)$ vanishes. But since the cofactor

D_{12} for any of the $\binom{n}{k}$ aspects cannot be negative the supposition that $D_{12}^k(G_1, G_2, \dots, G_n) = 0$ conversely implies that all those $\binom{n}{k}$ cofactors vanish.

The basic topological relation between the form of the transfer function of a 3TRN and the geometry of the network, whose somewhat lengthy proof is postponed to Section 10.2, may be formulated thus:

Theorem (7).—The transfer function of a 3TRN vanishes for each combination of k out of n conductances G_1, G_2, \dots, G_n short-circuited and the other $(n - k)$ open-circuited, being not identically zero, if, and only if, for each transfer path joining the input terminal with the output terminal one of two alternatives occurs, namely

- (a) It contains at least $(k + 1)$ conductances G_i .
- (b) Between the nodes through which such a transfer path passes and the reference node there exist at least $(n - k + 1)$ shunt paths,* composed exclusively of the conductances G_i , no G_i being common to two of them.

Moreover, the transfer paths joining the input and output terminals must all belong to the same family of paths, (a), or (b).

* Two shunt paths (bridges) are considered different if they differ in at least one element. Cf. the definition of a transfer path.

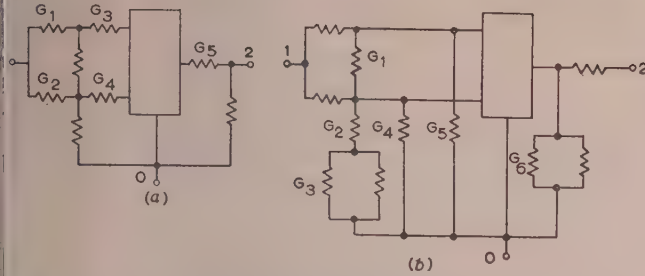


Fig. 4.—Two examples of a network with $D_{12}^k(G_1, G_2, \dots, G_n) = 0$.
(a) With $k+1 = 3$ series conductances G_i in each transfer path.
(b) With $n-k+1 = 3$ G_i -shunts joining the nodes of each transfer path with the reference node.

Fig. 4(a) shows an example of the alternative (a) with $n = 5$, $k = 2$, i.e. with at least three (here exactly three) out of five conductances G_i lying in each transfer path joining 1 and 2. Fig. 4(b), on the other hand, illustrates the alternative (b) for $n = 6$ and $k = 4$, i.e. $(n-k+1) = 3$. Here each transfer path between 1 and 2 is bridged to the reference node by at least three shunt paths, as defined in Theorem (7). For instance, for the upper path we may count the bridges G_5, G_6 and (G_1, G_2, G_3) or (G_1, G_4) . The last two bridges cannot be counted together since they have the conductance G_1 in common.

Taking into account the complementary transfer function, just as in the case of Theorem (5) we obtain

Theorem (8).—If the transfer function of a 3TRN is equal to unity for each combination of k out of n conductances G_1, G_2, \dots, G_n short-circuited and the other $(n-k)$ open-circuited, but is not identically unity, then the statement of Theorem (7) holds with the roles of the nodes 1 and 0 interchanged.

Now suppose that $D_{12}^k(G_1, G_2, \dots, G_n) = 0$ and that the alternative (a) of Theorem (7) holds. Evidently in this case $D_{12}^k(G_1, G_2, \dots, G_n) = 0$ for each $o \leq i \leq k$. If, on the other hand, the alternative (b) is valid, the same equality holds for each i with $k \leq i \leq n$. Thus, as a direct consequence of Theorems (7) and (8), we have

Theorem (9).—If $D_{12}^k(G_1, G_2, \dots, G_n) = 0$ [or $D_{12}^k(G_1, G_2, \dots, G_n) = D_{11}^k(G_1, G_2, \dots, G_k)$ for some k ($0 \leq k \leq n$)] then $D_{12}^k(G_1, G_2, \dots, G_n) = 0$ [or $D_{12}^k(G_1, G_2, \dots, G_n) = D_{11}^k(G_1, G_2, \dots, G_n)$] for each i with:
(a) $0 \leq i \leq k$
or
(b) $k \leq i \leq n$

(5) THE TRANSFER FUNCTION OF RC NETWORKS

Suppose now that, in addition to an arbitrary number of conductances, a network contains also n capacitances C_1, C_2, \dots, C_n . Evidently we can expand the cofactors D_{gh} of the admittance matrix of this network about the values of the susceptances pC_1, pC_2, \dots, pC_n in exactly the same way as before in the case of variables G_i . Thus we have identically

$$D_{gh}(pC_1, pC_2, \dots, pC_n) \equiv \sum_{k=0}^n D_{gh}^k(pC_1, pC_2, \dots, pC_n) \quad (10)$$

The symbol $D_{gh}^k(pC_1, pC_2, \dots, pC_n)$ denotes, as before, the sum of all $\binom{n}{k}$ products of k out of the n susceptances pC_i , each product multiplied by the cofactor D_{gh} taken for these k capacitances short-circuited and the other $n-k$ open-circuited.*

Since in all terms forming the D_{gh}^k the complex variable p appears in the same degree k we have

$$D_{gh}^k(pC_1, pC_2, \dots, pC_n) = p^k D_{gh}^k(C_1, C_2, \dots, C_n) \quad (11)$$

* For special cases cf. Section 10.1.

and

$$D_{gh}(pC_1, pC_2, \dots, pC_n) \equiv \sum_{k=0}^n p^k D_{gh}^k(C_1, C_2, \dots, C_n) \quad (12)$$

Note that all cofactors which take part in this expansion are cofactors of purely real matrices, which correspond to various aspects of the network. Taking this into account and using Theorems (1) and (2) we have

Theorem (10).—All cofactors D_{ik} of the admittance matrix of an RC network are polynomials in p with non-negative coefficients.

Theorem (11).—All coefficients of the polynomial expansion (in p) of D_{il} are not smaller than the corresponding coefficients of D_{ik} for each i and k ($i, k = 1, 2, \dots, n$).

These theorems have been proved by Fialkow and Gerst.⁶

Applying eqn. (12), the transfer function of a 3-terminal RC network with n capacitances (or shortly 3TRnCN) may be expanded about the values of all susceptances in the form

$$T(pC_1, pC_2, \dots, pC_n) \equiv \frac{\sum_{k=0}^n p^k D_{12}^k(C_1, C_2, \dots, C_n)}{\sum_{k=0}^n p^k D_{11}^k(C_1, C_2, \dots, C_n)} \quad (13)$$

Consider for a moment the first and the last terms in the numerator and the denominator of eqn. (13) only. For $k=0$ and $k=n$ the symbol D_{gh}^k contains only one term, and we have

$$T(pC_1, pC_2, \dots, pC_n) \equiv \frac{p^n C_1 C_2 \dots C_n D_{12}(s, s, \dots, s) + \dots + D_{12}(o, o, \dots, o)}{p^n C_1 C_2 \dots C_n D_{11}(s, s, \dots, s) + \dots + D_{11}(o, o, \dots, o)} \quad (14)$$

If $D_{12}(s, s, \dots, s) \neq 0$, the product of all zeros $p_i' (i = 1, 2, \dots, n)$ of the transfer function, eqn. (14), is evidently $D_{12}(o, o, \dots, o)/C_1 C_2 \dots C_n D_{12}(s, s, \dots, s)$, whereas the product of all its poles $p_i (i = 1, 2, \dots, n)$ equals $D_{11}(o, o, \dots, o)/C_1 C_2 \dots C_n D_{11}(s, s, \dots, s)$. Putting the first quotient over the second one we have

$$\frac{\prod_{i=1}^n p_i'}{\prod_{i=1}^n p_i} = \frac{D_{12}(o, o, \dots, o)}{C_1 C_2 \dots C_n D_{12}(s, s, \dots, s)} \cdot \frac{D_{11}(o, o, \dots, o)}{C_1 C_2 \dots C_n D_{11}(s, s, \dots, s)} = \frac{T(o, o, \dots, o)}{T(s, s, \dots, s)} \quad (15)$$

or in words,

Theorem (12).—The ratio of the product of all zeros of the transfer function of a 3TRnCN to the product of all its poles is equal to the ratio of the transfer function evaluated for the case of all capacitors open-circuited to the transfer function for the case of all capacitors short-circuited, provided that in the latter case this transfer function does not vanish.*

Applying Theorem (9) to the case of a network with n capacitors, we immediately have

Theorem (13).—If in the expansion, eqn. (13), of the transfer function of a 3TRnCN, the term with p^k is missing from the numerator, all terms of degree less than k , or all terms of degree greater than k , are also missing.

The two cofactors D_{12} and D_{11} may have common factors, and the order of the rational function† $T = D_{12}/D_{11}$ may then be lowered. What is preserved in the reduced transfer function is the difference of the highest and the difference of the lowest degrees of non-vanishing terms in the numerator and the

* For $n = 1$ this yields the familiar equation⁸

$$\frac{p_1}{p_1} = \frac{T(o)}{T(s)}$$

† The order of a rational function equals the greater of the degrees of the numerator and denominator.

denominator which must be the same as in the original form, eqn. (13). Let W and w denote the highest and the lowest degrees of non-vanishing terms in the denominator and let V and v have the same meaning for the numerator. By Theorems (10) and (11) we have $w \leq v \leq V \leq W$. As can be seen, if one of the differences $f = v - w$, or $h = W - V$ is not zero, Theorem (7) yields a useful indication on the geometry of such a network, namely if $f \neq 0$ each transfer path joining the input and output terminals contains at least f capacitors. But, since this difference is not $f + 1$, at least one transfer path must contain exactly f capacitors. On the other hand, if the difference of the highest degrees, h , is not zero, it is necessary to consider an arbitrary transfer path joining the input and output terminals. The nodes through which it passes may be, in general, joined to the reference node by means of a number of shunt paths, of which at least h must be composed exclusively of capacitors (with no capacitor common to two of them). Evidently, as in the previous case, for at least one transfer path the number of these shunts must be exactly h .

The complementary case to Theorem (13) may be formulated thus:

Theorem (14).—If in the expansion, eqn. (13), of the transfer function of a 3TRnCN the coefficients of p^k are equal in the numerator and the denominator, this property holds for the coefficients of all p^i with (a) $0 \leq i \leq k$, or (b) $k \leq i \leq n$.

As can be seen, the structure of such a network is analogous to that described above with the roles of the input and reference nodes interchanged.

(6) THE GAIN LIMIT OF THE RC LADDER TRANSFER FUNCTION

The upper limit for each coefficient in the numerator of the non-reduced expansion, eqn. (13), of the transfer function of a 3TRCN is the value of the corresponding coefficient in the denominator. But this limit may be considerably lower if there are some terms missing in the numerator.

At this stage it is desirable to prove some auxiliary theorems concerning the roots of the determinant of the admittance matrix of an RC network. The first of these is

Theorem (15).—The roots of the determinant of an RC network derived from the original network by short-circuiting or open-circuiting some capacitance are interlaced with the roots of the original network. The first and last points in the sequence of the interlaced roots are those of the original network.

In order to prove this, let C_1, C_2, \dots, C_n be the network capacitances and let $D(pC_1, pC_2, \dots, pC_n)$, $D(s, pC_2, \dots, pC_n)$ and $D(o, pC_2, \dots, pC_n)$ be the determinant of the original network, the determinant corresponding to short-circuiting C_1 and that with C_1 open-circuited, respectively. Since the admittance between two nodes of a network may be obtained by dividing the determinant of the original network by the determinant of the network with both those nodes short-circuited (Section 10.3), the admittance Y measured at the terminals of the capacitance C_1 is

$$Y = \frac{D(pC_1, pC_2, \dots, pC_n)}{D(s, pC_2, \dots, pC_n)} \quad (16)$$

Being an RC admittance, Y has its zeros real, negative, simple and interlaced with its poles,¹² the nearest to the origin and the furthest out (except for a pole at infinity) being zeros. Thus at both ends of the sequence of interlaced roots of the numerator and the denominator we find those of the numerator which proves Theorem (15) for the case of the short-circuit. Suppose now one end of the capacitor C_1 to be disconnected from the network and the admittance \bar{Y} measured between this terminal

and the node to which it was formerly connected. Analogously to the previous case we have:

$$\bar{Y} = \frac{pC_1 D(o, pC_2, pC_3, \dots, pC_n)}{D(pC_1, pC_2, \dots, pC_n)} \quad (17)$$

The numerator, which corresponds to the state of the network before the two terminals of \bar{Y} are short-circuited, is now equal to the product of pC_1 and $D(o, pC_2, \dots, pC_n)$, since the capacitor C_1 is hinged at one of its terminals to the network derived from the original one by putting $C_1 = o$. \bar{Y} has a zero at the origin which is not a zero of $D(o, pC_2, \dots, pC_n)$. Leaving out this zero there remains the sequence of interlaced roots of $D(pC_1, pC_2, \dots, pC_n)$ and $D(o, pC_2, \dots, pC_n)$ having at both ends the zeros of the former determinant, which proves Theorem (15) in this case also.

Let the roots of the determinant of an RC network be denoted by $-p_1, -p_2, \dots, -p_n$ with $o < p_1 < p_2 < \dots < p_n$. Theorem (15) leads at once to

Theorem (16).—If we short-circuit some capacitor of an RC network, the absolute value of the sum or the product of all roots of the determinant of each of the new networks cannot be smaller than $\sum_{i=1}^{n-1} p_i$ or $\prod_{i=1}^{n-1} p_i$, nor exceed $\sum_{i=2}^n p_i$ or $\prod_{i=2}^n p_i$ respectively.

The proof follows from the fact that, if for each root of the new determinant the nearest root of the original determinant with smaller absolute value be substituted, the result is the sum or the product of the smallest $n - 1$ numbers of the sequence p_1, p_2, \dots, p_n , whereas, when the substituting root is greater, the result is the sum or the product of the greatest $n - 1$ numbers of that sequence.

Assume now that in the non-reduced (except for a possible reduction by a numerical factor) expansion form, eqn. (13), of the transfer function of a 3TRnCN only one term in the numerator, say with p^k , is different from zero and that the denominator is a polynomial of n th degree with n different and negative roots $0 > -p_1 > -p_2 > \dots > -p_n$. In this case p^k is both the least and the greatest power which appears in the numerator of the transfer function. This implies that each transfer path between the input and output terminals must simultaneously fulfil two requirements. It must contain at least k out of the n capacitors, and between its nodes and the reference node we must be able to identify at least $(n - k)$ shunt paths composed exclusively of different capacitors. Since the network contains only n capacitors, each of these $(n - k)$ shunts consists simply of one capacitor and the set of k series and $(n - k)$ shunt capacitors is common to all transfer paths between 1 and 2. The condition that the transfer function should be of n th degree implies that it cannot be achieved by a network with less than n capacitors. This restriction excludes, e.g., such cases as parallel or series connection of two capacitors without a branch point between them. The structure actually represents a ladder generalized in such a manner that it may incorporate also some

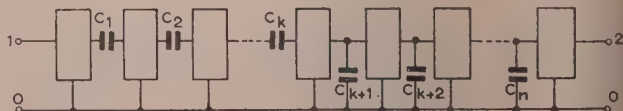


Fig. 5.—A generalized ladder network whose transfer function has a numerator with only one term different from zero.

3TRN as illustrated in Fig. 5 (the order of stages may of course be arbitrary).

Evidently only one aspect of such a network gives the transfer function different from zero. This unique aspect corresponds to short-circuiting all series and open-circuiting all shunt capacitors.

Without loss of generality the series capacitances may be denoted by C_1, C_2, \dots, C_k and the shunt ones by C_{k+1}, \dots, C_n . With this notation we have for the transfer function

$$\begin{aligned} r &= \frac{C_1 C_2 \dots C_k D_{12}(s, s, \dots, s, o, o, \dots, o) p^k}{C_1 C_2 \dots C_n D_{11}(s, s, \dots, s) p^n + \dots + D_{11}(o, o, \dots, o)} \\ &= \frac{K p^k}{\prod_{i=1}^n (p + p_i)} \quad (18) \end{aligned}$$

Applying (5.2) we have:

$$\kappa = \frac{D_{12}(s, s, \dots, s, o, \dots, o)}{C_{k+1} \dots C_n D_{11}(s, s, \dots, s)} \leq \frac{D_{11}(s, \dots, s, o, \dots, o)}{C_{k+1} \dots C_n D_{11}(s, s, \dots, s)} \quad (19)$$

Consider now the determinant $D_{11}(s, \dots, s, p C_{k+1}, \dots, p C_n)$ with the first k capacitors short-circuited. Its expansion about the values of the remaining $(n - k)$ susceptances is

$$\begin{aligned} D_{11}(s, \dots, s, p C_{k+1}, \dots, p C_n) \\ \equiv C_{k+1} \dots C_n D_{11}(s, \dots, s, s, \dots, s) p^{n-k} + \dots + \\ D_{11}(s, \dots, s, o, \dots, o) \quad (20) \end{aligned}$$

The quotient on the right side of eqn. (19) may be identified as the absolute value of the product of the $(n - k)$ roots of eqn. (20). Now remove the short-circuit from the capacitor C_k . According to Theorem (16) applied to the network with the terminals 1 and 0 short-circuited (D_{11}) the absolute value of the product of the roots of $D_{11}(s, \dots, s, p C_{k+1}, \dots, p C_n)$ is not greater than that of the greatest $(n - k)$ roots of $D_{11}(s, \dots, s, p C_k, C_{k+1}, \dots, p C_n)$. Proceeding further in the same manner, remove the short-circuits one after another from all capacitors C_{k-1}, \dots, C_1 and finally the original state of the network is reached with all capacitors recovered. Since the roots of $D_{11}(p C_1, p C_2, \dots, p C_n)$ are the poles of the transfer function, eqn. (18), the gain factor K cannot in this case exceed the product of the $n - k$ greatest poles of the transfer function, i.e. $\prod_{i=k+1}^n p_i$.

This limit is the lowest upper limit for K since, as is evident, this value can be approached as closely as desired in the ladder structure given in Fig. 6 if $C_i R_i$ is selected to equal $1/p_i$ and the

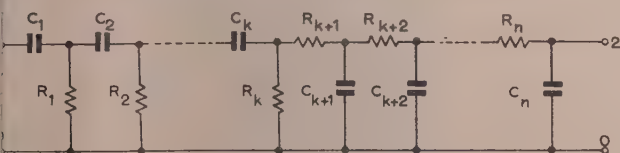


Fig. 6.—A ladder network whose transfer function has the gain limit equal to $\prod_{i=k+1}^n p_i$.

loading effect is diminished on every stage by increasing sufficiently the impedance level of the following stage. Since the ladder is composed of n capacitors and n resistors it contains the minimum number of elements necessary for realization of the transfer function, eqn. (18). Thus we have

Theorem (17).—The most general 3-terminal RC network with the transfer function of the type eqn. (18) having only n capacitors has a structure as exemplified by Fig. 5 (with an arbitrary order of stages). The upper limit for the gain attainable in such a network is equal to $\prod_{i=k+1}^n p_i$ and it may be approached arbitrarily closely by the ladder network of Fig. 6 built from the minimum number, $2n$, of elements.

(7) CONCLUSION

The method presented enables some properties of RC networks to be deduced from the properties of pure R networks by applying topological arguments. In the paper it has been applied chiefly to the transfer function of a 3-terminal RC network when working between zero source and infinite load impedances.

This is by no means the only field where the method may be used, nor should the limitation to the unbalanced structure be regarded as its natural one. Some other applications, e.g. to the zeros of the RC network transfer function and the selective properties of such networks, will be the subject of a further publication.

(8) ACKNOWLEDGMENT

The author wishes to express his gratitude to Dr. A. Talbot, Imperial College of Science and Technology, University of London, for his continued interest in this work, and to the Scientific Department, Ministry of Defence, Israel, under whose auspices the research was done.

(9) REFERENCES

- (1) SCOTT, R. E., and BLANCHARD, R. L.: 'An Iterative Method for Network Synthesis', Transactions of the Institute of Radio Engineers Professional Group on Circuit Theory, December, 1953, PGCT-2, p. 19.
- (2) CARTER, C. W., JR.: 'Graphic Representation of the Impedance of Networks containing Resistances and Reactances', *Bell System Technical Journal*, 1925, 4, p. 387.
- (3) FELDTKELLER, R.: 'Einführung in die Vierpoltheorie der elektrischen Nachrichtentechnik' (Hirschel, Leipzig, 1937), p. 25.
- (4) REZA, F. M., and LEWIS, P. M.: 'A Note on the Transfer Voltage Ratio of Passive RLC Networks', *Proceedings of the Institute of Radio Engineers*, 1954, 42, p. 1452.
- (5) TALBOT, A.: 'Some Fundamental Properties of Networks without Mutual Inductance', *Proceedings I.E.E.*, Monograph No. 118 R, January, 1955 (102 C, p. 168).
- (6) FIALKOW, A., and GERST, I.: 'The Transfer Function of General Two-Terminal-Pair RC Networks', *Quarterly of Applied Mathematics*, 1952, 10, p. 113.
- (7) BODE, H. W.: 'Network Analysis and Feedback Amplifier Design' (Van Nostrand, New York, 1945), p. 91.
- (8) MACRAE, D., JR.: 'Low-Frequency Amplifiers with Stabilized Gain', M.I.T. Radiation Laboratory Series, 1948, 18, p. 344.
- (9) PERCIVAL, W. S.: 'The Solution of Passive Electrical Networks by means of Mathematical Trees', *Proceedings I.E.E.*, Paper No. 1492 R, May, 1953 (100, Part III, p. 143).
- (10) PERCIVAL, W. S.: 'Improved Matrix and Determinant Methods for Solving Networks', *ibid.*, Monograph No. 96 R, April, 1954 (101, Part IV, p. 258).
- (11) SHEKEL, J.: 'Two Network Theorems concerning Change of Voltage Reference Terminal', *Proceedings of the Institute of Radio Engineers*, 1954, 42, p. 1125.
- (12) GUILLEMIN, E. A.: 'Communication Networks' (Wiley, New York, 1935), 2, pp. 211–216.

(10) APPENDICES

(10.1) Proof of Eqn. (4)

In the case when G_1 connects some node, other than 1 and 2, with the reference node, and thus appears only on one place in the admittance matrix, eqn. (4) is well known.⁷

In the general case, suppose that G_1 connects some two nodes, say i and k , one of which at least, say k , is different from 1 and 2. G_1 then appears in the admittance matrix on four places (ii) , (ik) , (ki) and (kk) . Let the k th row and column be added to the i th row and column respectively (which does not alter the value of the determinant). After this is done G_1 appears only on the place (kk) , and the coefficients of G_1 in the network determinant and in both cofactors D_{11} and D_{12} may be obtained by deleting the k th row and column of the transformed determinant or cofactor. But the combined operation of adding the k th row and column to the i th row and column followed by deleting the k th row and column is equivalent to short-circuiting the nodes i and k , which justifies eqn. (4).

Special attention must be paid to the cases when G_1 connects two out of the three terminals 0, 1 and 2, since after short-circuiting G_1 one of these terminals disappears. To ensure that eqn. (4) is comprehensive some conventions must be adopted here.

(a) If G_1 connects 0 and 1, $D_{11}(G_1)$ and $D_{12}(G_1)$ are independent of G_1 ; therefore we write in this case $D_{11}(s) = D_{12}(s) = 0$.

(b) If G_1 connects 0 and 2, D_{12} is independent of G_1 and therefore we put $D_{12}(s) = 0$.

(c) If G_1 connects 1 and 2, since $\lim_{G_1 \rightarrow \infty} T = 1$, we write $D_{12}(s) = D_{11}(s)$, where $D_{11}(s)$ is the value of D_{11} for the network obtained by identifying terminal 2 with terminal 1 (G_1 short-circuited).

(10.2) Proof of Theorem (7)

The conditions of Theorem (7) are sufficient. To prove this, take any combination of k conductances G_i short-circuited and the remaining $(n - k)$ open-circuited. If some transfer path belongs to the (a) family then passing through at least $(k + 1)$ G_i conductances it must be interrupted. If a transfer path belongs to the (b) family, the $(n - k)$ open-circuited G_i can open at most $(n - k)$ out of the $(n - k + 1)$ bridges to the reference node, and therefore such a path must be short-circuited to the reference. Thus for any such combination we have $D_{12} = 0$, which implies $D_{12}^k(G_1, G_2, \dots, G_n) = 0$.

To prove the necessity, suppose that there exists a transfer path between 1 and 2 which does not fulfil the conditions (a) and (b). Such transfer path passes at most through k conductances G_i and has at most $(n - k)$ G_i -bridges to the reference node. Evidently there exists such a combination of k short-circuited and $(n - k)$ open-circuited G_i which closes all series G_i conductances of this transfer path and opens all bridges to the reference node, maintaining the transfer between 1 and 2 effective. Thus in this case $D_{12}^k(G_1, G_2, \dots, G_n) \neq 0$, which proves the necessity of the conditions.

To prove that all the paths belong to the same family, (a) or (b), when $D_{12}^k(G_1, G_2, \dots, G_n) = 0$, suppose to the contrary that there exists one path belonging to the family (a) and another (or the same) path belonging to the family (b). Then the former path contains at least $(k + 1)$ of the conductances G_i , none of which is joined to the reference node, whereas the latter path is shunted by a number of paths containing at least $(n - k + 1)$ of the conductances G_i joined to the reference node, and therefore distinct from the former $(k + 1)$ conductances, as well as from one another. Thus there must exist in the network $(k + 1) + (n - k + 1) = (n + 2)$ conductances G_i contrary to the assumption.

(10.3) Admittance between Two Nodes of a Network

Although the proposition quoted in proof of Theorem (15) may be found in the literature,⁹ a short proof of it will be given here.

If one of the nodes between which the admittance is measured is the reference node 0 and the other is, say, the k th one, we may use the well-known formula

$$Y = \frac{D}{D_{kk}}$$

The minor D_{kk} , obtained from the original determinant by deleting the k th row and the k th column is the determinant of the network with the mode k identified with the reference node 0.

Now the node determinant of the network is invariant with respect to changes of the reference node.^{10,11} Thus if a new reference node is chosen, other than 0 and k , the new network determinant D' equals D , and the determinant of the network with the nodes k and 0 identified with respect to the new reference equals D_{kk} . This proves the proposition.

INTEGRAL CONTROL WITH TORQUE LIMITATION

By J. C. WEST, Ph.D., Associate Member, and M. J. SOMERVILLE, B.Sc., Graduate.

(The paper was first received 20th October, 1955, and in revised form 8th February, 1956. It was published as an INSTITUTION MONOGRAPH in June, 1956.)

SUMMARY

The behaviour of a basic remote-position-control system with velocity-feedback stabilization and integral-of-error control is investigated, taking torque limitation of the servo motor into account. It is shown that the transient response deteriorates as the magnitude of the input signal is increased and becomes unstable above a critical value. Both phase-plane analysis and describing-function techniques are employed. The experimental behaviour to random input signals is observed.

The effect of other non-linear elements added to eliminate the possibility of amplitude-dependent instability is discussed.

LIST OF PRINCIPAL SYMBOLS

- a = Value of acceleration when torque saturation ensues.
 v_c = Control signal.
 v_m = Saturation level of control signal.
 F_v = Torque produced by viscous friction per unit angular speed.
 $F(\omega)$ = Frequency-response transfer-function for the broken loop.
 f = Frequency.
 h = Magnitude of initial input step.
 $h_c = \frac{2}{3}a\tau T_i$ = Critical magnitude of input step to make system unstable.
 $H = x/h_c$ = A change of variable measuring position in terms of the critical step magnitude.
 I, I_0 = Signal from the integrator. The suffix 0 refers to initial value at $t = 0$.
 J = Moment of inertia of rotating parts.
 $K = y_{max}/y_c$ = Ratio of the maximum limited velocity to the critical velocity.
 m = Gain.
 $N(h)$ = Describing-function in terms of amplitude h for a given non-linearity.
 p = Heaviside operator.
 t = Time, sec.
 τ = Time-constant of velocity damping.
 τ_v = Time-constant of damping due to friction.
 T_i = Integrator time factor.
 $u = y/y_c$ = A normalized variable representing velocity in terms of the critical velocity.
 x = Output position.
 z = Output acceleration.
 x_0, y_0, z_0 = Initial values of x, y, z at $t = 0$.
 $y_c = a\sqrt{(3\tau T_i)}$ = Critical velocity attained for the critical input step h_c .
 y_{max} = Maximum velocity set by velocity-limiting network.
 $\alpha = \tau\omega_0$ = Damping factor.
 $\beta = \sqrt{(3T_i/\tau)}$ = A useful coefficient of the ratio of the integrator time factor and the damping time-constant.

$\delta = \alpha T_i \omega_0 = A$ design factor to determine the value of T_i satisfactory for linear operation.

θ_i, θ_o = Input and output positions.

$T_n = \omega_0 t / \alpha \beta$ = Normalized time to allow variation of several parameters.

$\omega = 2\pi f$.

$\omega_0/2\pi$ = Undamped natural frequency of oscillation of the basic linear system.

(1) INTRODUCTION

The addition of integral of error to the control signal of a linear servo-mechanism has the desirable effect of reducing considerably the steady-state and low-frequency errors. It is well known, however,^{1,2} that for such systems the sudden occurrence of a large error signal which temporarily saturates some part of the system can produce a low-frequency oscillation of excessively large amplitude. This is due to the action of the integrator, and mainly for this reason 'integral of error' control has not been adopted to the extent that it merits. In systems where it has been included, arrangements are often made to switch out the integrator at large errors, as in the initial running into alignment, or, in two-channel coarse/fine systems, the integrator may be placed only in the fine-error channel.

It is found that satisfactory performance can be obtained if an integrator circuit is employed which has the ability to cease integrating when the integrated signal rises above a predetermined level. This level becomes a design parameter of the system and is related to the control-signal level at which saturation of the forward chain occurs.

The undesirable action of the integrator is best shown by the consideration of step-function responses, and for this reason the phase-plane method of analysis has been employed. This is normally restricted to second-order equations, whereas the simplest integral control has a third-order characteristic equation. It is found possible to utilize the phase-plane technique provided that all the trajectories start from a state of rest, i.e. zero initial velocities and accelerations.

In order to discuss the stability of these systems in the presence of saturation it has been found useful to use the frequency-response approach and the describing-function technique. It has not, however, been found practical to analyse the system with two non-linearities present at the same time, i.e. saturation of the amplifier or servo motor and the deliberate amplitude-limited integrator.

The above two methods of analysis show the improvement that can be achieved in stability and transient response by use of the limited integrator. They fail to show the behaviour for the normal type of continuously varying signal encountered for the major portion of the working period of such a system. Results have been obtained to show the effect of limited integral control on the ability of the system to follow a stochastic signal with fixed amplitude-probability distribution and power-frequency spectrum.

Experimental work was based on a servo model using a

Correspondence on Monographs is invited for consideration with a view to publication.
 Dr. West and Mr. Somerville are in the Electrical Engineering Department, University of Manchester.

Type 74 split-field d.c. motor, and an all-electronic simulator of the model working at a natural frequency 200 times as great as that of the model.

(2) THE BASIC SYSTEM

The simplest position-control system employing integral-of-error and stabilized by direct velocity feedback is shown schematically in Fig. 1. In the linear regime the torque (and in

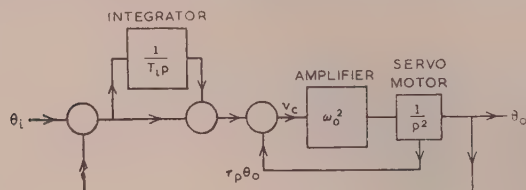


Fig. 1.—Basic control system.

the absence of friction the acceleration also) is proportional to the control signal v_c . This is composed of the addition of three terms, one directly proportional to error, one to the integral of error and the third to the output velocity term which introduces a stabilizing influence; thus

$$v_c = (\theta_i - \theta_o) + \frac{1}{T_i} \int_0^t (\theta_i - \theta_o) dt - \frac{\tau_p d\theta_o}{dt}$$

$$= \left(\frac{1 + T_i p}{T_i p} \right) (\theta_i - \theta_o) - \tau_p \dot{\theta}_o \quad (1)$$

The torque produced per unit control signal is $J\omega_o^2$, which is a design constant of the system including the loop gain and the moment of inertia J of the rotating parts. The total torque is therefore

$$J\omega_o^2 v_c \quad (2)$$

$$= J\ddot{\theta}_0 + F_v \dot{\theta}_0 \quad (3)$$

where F_v is the torque produced by viscous friction and will eventually be assumed negligible.

From the first three equations the overall characteristic equation for the linear system is obtained in this form

$$\left[\frac{T_i}{\omega_o^2} p^3 + (\tau + \tau_v) T_i p^2 + T_i p + 1 \right] \theta_o = (1 + T_i p) \theta_i \quad (4)$$

with $\tau_v = F_v/J$ for convenience.

(2.1) Stability of the Linear System

Without the integrator, i.e. $T_i = \infty$, the characteristic equation is

$$\left[\frac{p^2}{\omega_o^2} + (\tau + \tau_v) p + 1 \right] \theta_o = \theta_i$$

with ω_o the undamped natural frequency and $(\tau + \tau_v)$ the damping term. In practice τ would be arranged to give a reasonable response. $(\tau + \tau_v)\omega_o = \alpha$ may be called the damping factor, and $\alpha = 2$ corresponds to critical damping.

For positive damping, the system cannot be unstable. The addition of integral control and the determination of the required value of T_i would be achieved after the damping factor had been chosen.

The Hurwitz-Routh³ conditions for stability applied to the characteristic equation, eqn. (4), give⁴

$$(\tau + \tau_v) T_i \omega_o^2 > 1$$

$$T_i \omega_o > 1/\alpha \quad (5)$$

Let $T_i \omega_o = \delta/\alpha$. The necessary factor δ , by which $T_i \omega_o$ exceeds $1/\alpha$ to give a reasonable response, depends on the value of α itself; when $\alpha = 2$, $\delta = 12$ (and $T_i \omega_o = 6$) is a satisfactory value.

The basic linear stable system having been defined, it is now possible to consider the effects of torque saturation of the servo motor.

(2.2) Motion in the Saturated Regime

It is assumed that torque-limiting or saturation of the servo motor will occur when the control signal exceeds a certain value v_m , i.e. when $|v_c| > |v_m|$.

Under these conditions the total torque is a constant and is independent of the control signal whilst it still exceeds v_m .

If, for simplicity, the viscous-friction term is assumed negligible, the equation of motion under saturated conditions becomes

$$\frac{d^2 \theta_o}{dt^2} = \pm a = \pm \omega_o^2 v_m \quad (6)$$

the solution of which, with allowance for friction, is given in Reference 7.

If the system is initially at rest with zero velocity and acceleration and in the condition in which

$$\theta_i = \theta_o = -h$$

an application of an input step of magnitude h immediately takes θ_i from position $-h$ to the arbitrary zero position. The error, since the output cannot move instantaneously, jumps from zero to h .

It will be assumed that h is many times greater than v_m and hence the system will immediately become saturated and the ensuing motion will be governed by eqn. (6).

By integrating with respect to time, the output velocity and position may be obtained; thus

$$\frac{d\theta_o}{dt} = at \quad (7)$$

$$\theta_o = \frac{1}{2} at^2 - h \quad (8)$$

so that the error signal is

$$\theta_i - \theta_o = h - \frac{1}{2} at^2 \quad \text{for } t > 0$$

$$= 0 \quad \text{for } t < 0$$

Let the signal from the integrator unit be I ; then in general

$$I = \frac{1}{T_i} \int_0^t (\theta_i - \theta_o) dt + I_0 \quad (9)$$

where I_0 is the initial value, which for this particular case is zero (see also Section 9.2). Hence

$$I = \frac{1}{T_i} \int_0^t \left(h - \frac{1}{2} at^2 \right) dt = \frac{1}{T_i} \left(ht - \frac{1}{6} at^3 \right)$$

The control signal is, from eqn. (1),

$$v_c = (\theta_i - \theta_o) + I - \tau \frac{d\theta_o}{dt}$$

and the saturated motion continues until this signal has fallen to the value v_m . Further reduction brings the system into a linear operating condition.

Fig. 2 shows the various components of the control signal as functions of time. The velocity damping term $\tau d\theta_o/dt$ becomes τat and increases linearly with time. The line $(h - \tau at)$ is drawn

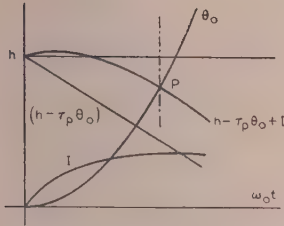


Fig. 2.—Determination of the first boundary.

for convenience. The curve $(h - \tau a t + I)$ is also drawn so that its distance from that of the output θ_0 gives the control signal v_c . If h is very much greater than v_m , the time at which the control signal reaches the value v_m is very nearly at, but slightly before, the intersection of these two curves.

Subsequent motion passes through the linear regime into the saturated condition of maximum deceleration. Thus, if the change-over from maximum acceleration to maximum deceleration occurs before the error $\theta_i - \theta_0$ has fallen to zero, the overshoot will be of less amplitude than h . If, however, the initial step is so great that the intersection, point P of Fig. 2, occurs after θ_0 has reached zero level (i.e. P is above the horizontal line h), the overshoot will be greater than h and the system will oscillate with ever increasing amplitude. The critical magnitude h_c of input step for which the output executes oscillations of equal amplitudes $\pm h_c$ occurs when P lies on the horizontal line h , i.e. when the error is zero.

If this occurs at time t_1 , we have

$$\theta_i - \theta_0 = h_c - \frac{1}{2} a t_1^2 = 0$$

$$t_1 = \sqrt{\frac{2h_c}{a}}$$

The velocity attained at this time t_1 is called y_c and is

$$y_c = \sqrt{(2ah_c)} = a t_1$$

The condition for point P to be on the horizontal line h is that

$$h - \frac{1}{2} a t_1^2 - \tau a t_1 + I = 0$$

at time t_1 , i.e.

$$I = \tau a t_1$$

$$h_c - \frac{1}{2} a t_1^2 = \tau T_i a$$

$$h_c - \frac{3}{2} h_c = \tau T_i a$$

$$h_c = \frac{3}{2} \tau T_i \omega_0^2 v_m \quad \dots \quad (10)$$

this can be written in the form

$$h_c = \frac{3}{2} \delta v_m$$

thus an applied step of the order $\frac{2}{3} \delta$ times the extent of the linear regime produces instability. For the particular system described in Section 2.1, $\alpha = 2$, $\delta = 12$ and $(h_c/v_m) = 18$.

(2.3) Phase-plane Analysis

In order to study the motion of the system for any arbitrary step h and to determine the regions of saturation and linear operation, it is preferable to obtain solutions independent of the time t by use of phase-plane analysis. The phase plane is a curve of output velocity $y \equiv d\theta_0/dt$ as ordinate against position, θ_0 , as abscissa.

The saturated motion given by eqn. (6) can be written

$$\frac{d^2 \theta_0}{dt^2} = \frac{dy}{dt} = \frac{dy}{dx} \frac{dx}{dt} = y \frac{dy}{dx} = \pm a$$

Hence, by direct integration with respect to x , the trajectory in the phase plane becomes

$$\int_{y_0}^y y dy = \pm \int_{x_0}^x a dx$$

$$y^2 - y_0^2 = \pm 2a(x - x_0)$$

where x_0 and y_0 are initial conditions.

For a step of magnitude h as in Section 2.2, the equation of motion becomes

$$y^2 = 2a(x + h)$$

This equation could have been derived alternatively from eqns. (7) and (8) by eliminating t .

This motion ensues all the time saturation is present, but the linear equation of operation (eqn. 4) predominates as soon as the control signal v_c has fallen to the value v_m . This change-over condition can be shown as a curve in the phase plane of such form that when the trajectory meets it the initial saturated mode of operation is terminated.

For the cases of interest when $h \gg v_m$ the effect of the passage through the linear regime is negligible and it is possible to neglect the width of the linear regime with very little loss in accuracy but with great simplification of the analysis. The change-over boundary in the phase plane can be considered as the position for which the control signal is zero and it can be assumed that acceleration changes from $+a$ to $-a$ on crossing the boundary.

The analysis is given in Section 9.1 to determine the equations of these boundaries, some of which are shown in Fig. 3 together with the trajectory for the critical step h_c , namely

$$y^2 = 2a(x + h_c)$$

which meets the change-over boundary at $x = 0$ and $y_c = \sqrt{(2ah_c)} = a\sqrt{(3\tau T_i)}$ from eqn. (10), as was shown in Section 2.2.

For analytical simplification the axes of the phase plane are normalized in terms of h_c and y_c and the new variables of output position and speed are

$$\begin{aligned} H &= x/h_c \\ u &= y/y_c \end{aligned} \quad \dots \quad (11)$$

An input step of h now becomes an input step of $H = h/h_c$.

The equations of the boundaries become (see Section 9)

$$H = \frac{2(u^2 - 1)u}{3au + \beta} \quad \dots \quad (12)$$

$$\text{where } \beta = \sqrt{\frac{3T_i}{\tau}}$$

and can be drawn by taking a series of values of u and calculating H .

This is shown in Fig. 3 for several values of β for the case in which v_m/h_c is negligibly small.

(2.4) Subsequent Boundaries with reference to the Linear Regime

Construction of subsequent boundaries may be obtained by extending the method employed for the first boundary. This method yields a cubic equation with parameters dependent upon the initial conditions at the previous boundary, and the determination of each point upon the new boundary involves the solution of a cubic equation.

A graphical method may be employed in the construction of all the boundaries, when the linear regime is assumed negligible; this is shown in Fig. 4 and described in Section 9.2. An extension of the method to include the linear regime is included in Section 9.3.

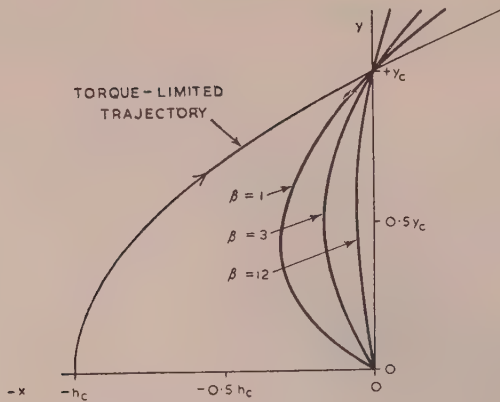


Fig. 3.—First boundaries as v_m/h_c approaches zero.

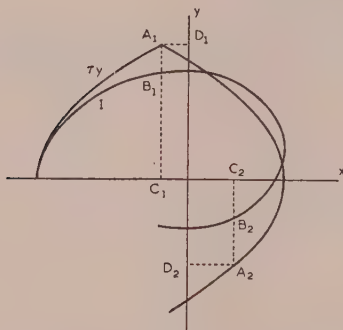
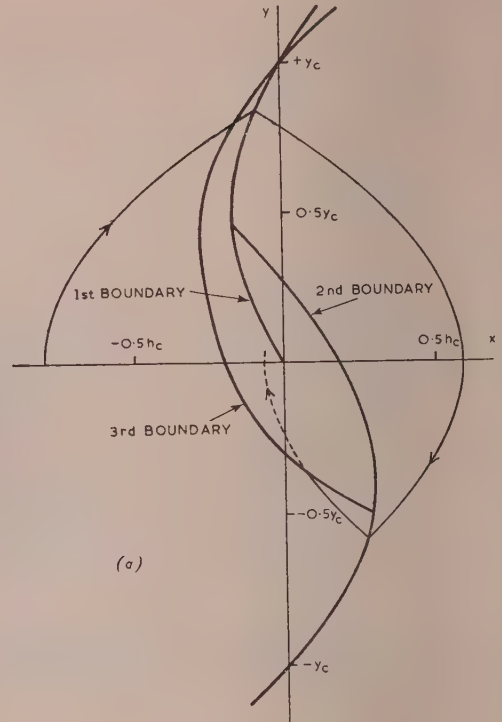


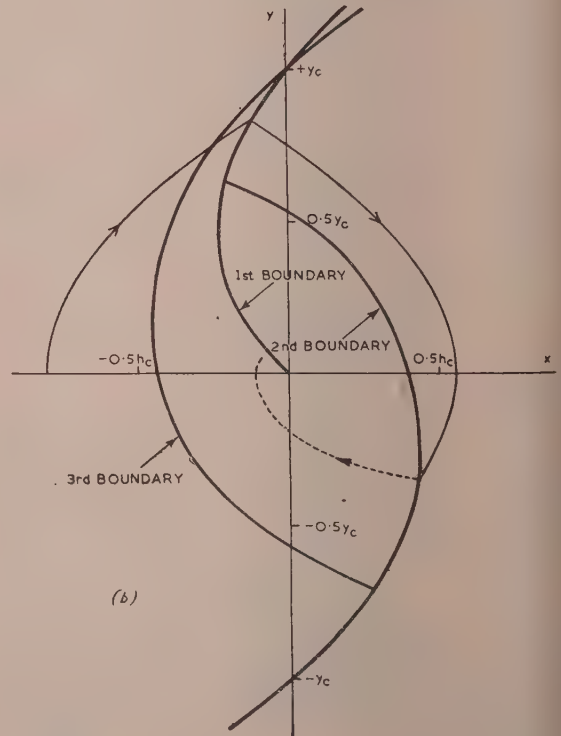
Fig. 4.—Graphical determination of boundaries.

The phase-plane boundaries shown in Figs. 5(a) and 5(b) for two different values of β have been constructed graphically by this method. The number of boundaries will be infinite because the value of I , which was zero at the instant of applying an input step, is finite and variable at subsequent crossings of the x axis. For $T_i = \infty$ the equation of motion becomes that for the second-order system, and all the boundaries coincide along a line of slope $dy/dx = -1/\tau$, tangential to the first boundary at the origin. Responses are respectively stable or unstable, according to whether the trajectory meets the first boundary before or after crossing the y axis, corresponding to torque reversal either before or after the error becomes zero. Therefore, for the systems considered, it is possible to obtain stability by ensuring that an even boundary does not cross the $-y$ axis or an odd boundary cross the $+y$ axis.

Consecutive boundaries intersect, so that the first boundary starts at the origin, the second boundary starts on the first boundary, and so on. Parabolic motion of the trajectory ensues from one boundary to the next, reversing torque each time a boundary is crossed. For a stable response, successive amplitudes of overshoot and undershoot are reduced until a boundary is reached at a point on it before the start of the next boundary, when the ensuing response becomes immediately unsaturated. Subsequent motion may take place entirely in the linear regime, although it is possible that the saturated region may again be entered, these boundaries not being shown. These sections of responses are shown in the diagrams by dotted lines. Time responses obtained from Figs. 5(a) and 5(b) are given in Figs. 8A and 9A.



(a)



(b)

Fig. 5.—Phase-plane trajectory for an input step of $0.8h_c$ showing the first three boundaries.

(a) $\beta = 3$.
(b) $\beta = \sqrt{3}$.

3) ADDITION OF NON-LINEAR ELEMENTS TO CORRECT FOR LARGE INPUT SIGNALS

Stabilization of the response to large input signals may be achieved by the introduction of a source of non-linearity to render the integrator inoperative for large error signals. Saturation-type limits applied either to the input or to the output of the integrator do not fulfil this function. Placed at the input, they limit the signal to the integrator, but this does not prevent the output signal becoming excessive. Output limits restrict the magnitude of the output for large input signals but do not alter the actual integrated voltage. In both cases the response must be such that a negative error shall exist long enough to remove the integrated charge accumulated during the positive-error regions. A method which has been found to give very effective stabilization is that illustrated in Fig. 6A, where the non-linear element

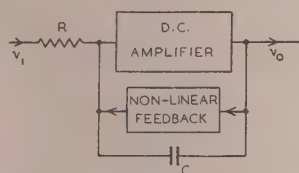


Fig. 6A.—Block diagram of limited integrator.

effectively short-circuits the integrating capacitor for voltages across it greater than $\pm V$. This arrangement also ensures that overloading of the integrator amplifier does not occur. The required non-linear feedback characteristic is that shown in Fig. 6B, which can be obtained by a pair of biased diodes. A further

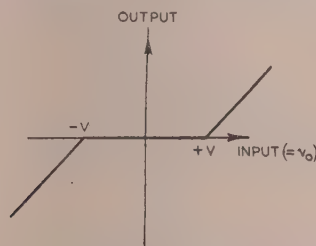


Fig. 6B.—Characteristic of the non-linear feedback element.

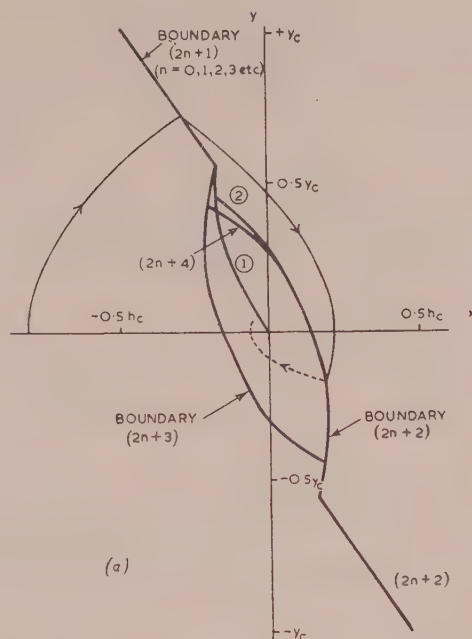
source of non-linearity which limits the sum of error plus integral of error, has been found to give extra stabilization to large input signals when used in conjunction with the integrator limiting already described. It has the effect of limiting the maximum velocity achieved during the response, and has been called velocity limiting.⁷ The step-function responses for the limited integrator with and without velocity limiting are derived by means of the phase-plane. The effect on overall stability of each of the various non-linearities is considered by the describing-function technique in Section 4.

(3.1) Limited Integrator

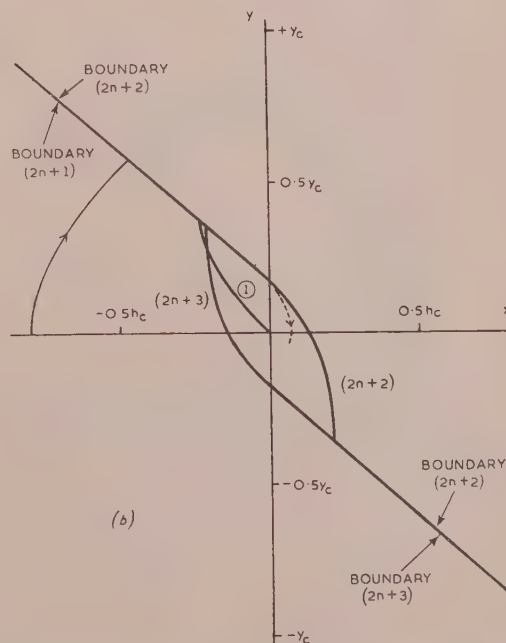
Boundaries for the step-function response are obtained by the graphical method described in Section 2.4. The curve of I against x is given by eqn. (18) (Section 9.2) until I reaches the maximum value $\pm V$, when this value is maintained until x changes sign. The boundaries in the phase-plane when $I = \pm V$ are given by $x + \tau y = \pm V$ for parabolic trajectories, so that

$$\frac{dy}{dx} = -\frac{1}{\tau} \text{ and } x = \pm V \text{ at } y = 0 \quad (13)$$

Figs. 7(a) and 7(b) show the boundaries for two values of β . A finite number of different boundaries is obtained, since the



(a)



(b)

Fig. 7.—Phase-plane trajectory for an input step of $0.8h_c$ showing the boundaries with a limited integrator.

- (a) $\beta = 3$; $V = 0.3V_{max} = 0.2h_c$.
 (b) $\beta = \sqrt{3}$; $V = (\sqrt{3}/10)V_{max} = 0.2h_c$.

value of I is reset to $\pm V$ each time limiting occurs in the integrator. The boundaries given by eqn. (13) are separated by $\pm V$ from the boundary for $T_i = \infty$, and responses to large signals are therefore very similar to those obtained for the corresponding second-order system.⁸ The maximum value for V which will just stabilize for large signals is such that the boundaries given

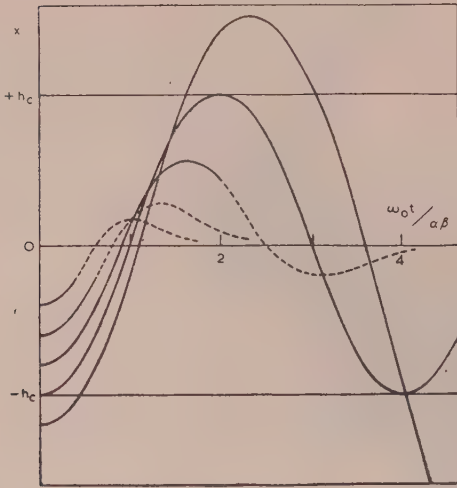


Fig. 8A.—Responses of the integral control system with torque saturation level $\pm \frac{2h_c}{\alpha^2\beta^2}$; $\beta = 3$.

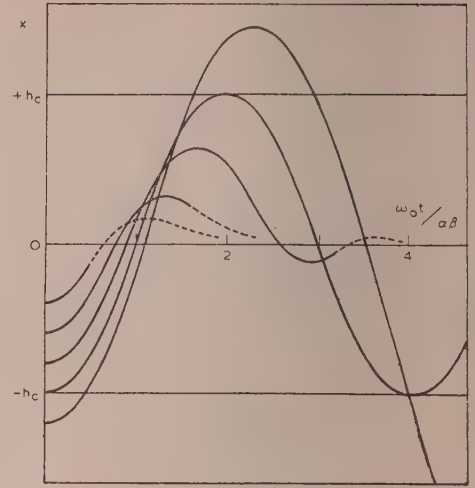


Fig. 9A.—Responses of the integral control system with torque saturation level $\pm \frac{2h_c}{\alpha^2\beta^2}$; $\beta = \sqrt{3}$.

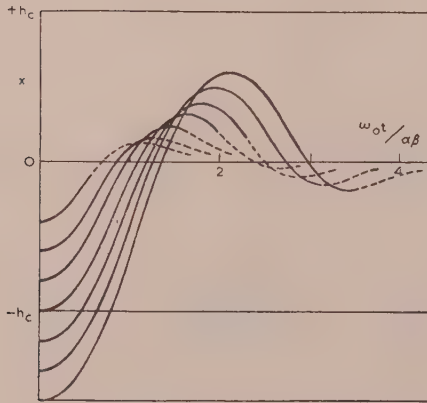


Fig. 8B.—Responses with the limited integrator, $\beta = 3$; $V = 0.3V_{max} = 0.2h_c$.

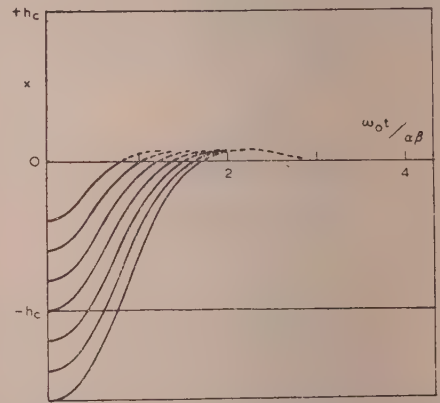


Fig. 9B.—Responses with the limited integrator, $\beta = \sqrt{3}$; $V = (\sqrt{3}/10)V_{max} = 0.2h_c$.

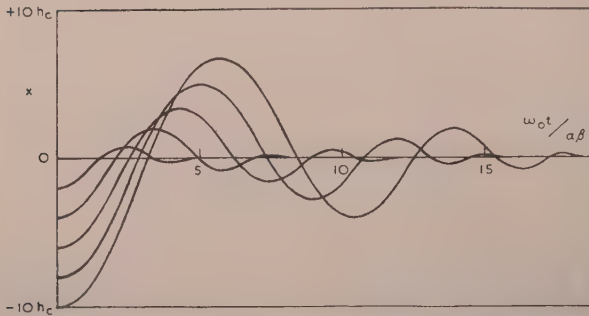


Fig. 8C.—Responses up to $\theta_i = 10h_c$ for the system with limited integrator, $\beta = 3$; $V = 0.3V_{max} = 0.2h_c$.

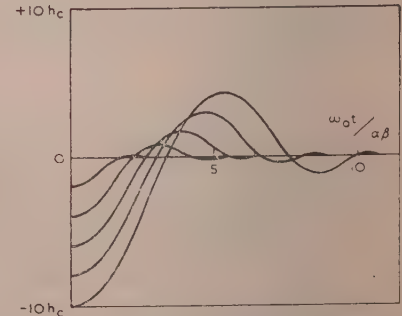


Fig. 9C.—Responses up to $\theta_i = 10h_c$ for the system with limited integrator, $\beta = \sqrt{3}$; $V = (\sqrt{3}/10)V_{max} = 0.2h_c$.

by eqn. (13) would pass through $\pm y_c$, thereby preventing the first boundary from crossing the $+y$ axis, namely

$$V_{max} = \tau y_c = \frac{2h_c}{\beta}$$

since

$$y_c = a\sqrt{3\tau T_i}$$

$$\beta = \sqrt{\frac{3T_i}{\tau}}$$

and

$$2h_c = 3a\tau T_i$$

It is possible for a trajectory on the phase plane of Fig. 7(b) to meet an odd boundary, given by eqn. (13), to reverse acceleration and to reach the consecutive (even) boundary before x has changed sign and the integrator output has thereby been reduced from its limited value $\pm V$. Consecutive odd and even boundaries will therefore partly coincide, their equations for the common region being identical [eqn. (13)]. Whether the slope of the phase-plane trajectory at its point of intersection with the odd boundary is less or greater than the modulus of the slope of the boundary ($+1/\tau$), will determine whether the even boundary will be reached immediately or after a section of parabolic trajectory. When a trajectory meets an even boundary in the manner outlined above, the ensuing trajectory will run along the even boundary until x changes sign and therefore obeys the equation $\dot{x} + \tau y = \pm V$. It is necessary to have this equation in terms of a time solution for construction of the time response; thus

$$-x \pm V = \tau \frac{dx}{dt}, \text{ whence } x \pm V = (x_0 \pm V)e^{-\omega_0 t/\alpha} \quad (14)$$

The values of V/V_m in Figs. 7(a) and 7(b) have been chosen to show the effect of maintaining both V and h_c fixed whilst varying the integrator time-constant T_i . The corresponding time responses are shown in Figs. 8b, 8c, 9b and 9c; Figs. 8a and 9a, for comparison, are the responses without the limitation on integral control.

(3.2) Velocity Limiting plus Integrator Limiting

The phase-plane boundaries of Fig. 10 are again constructed as in Section 3.1, and are similar to those for integrator limiting

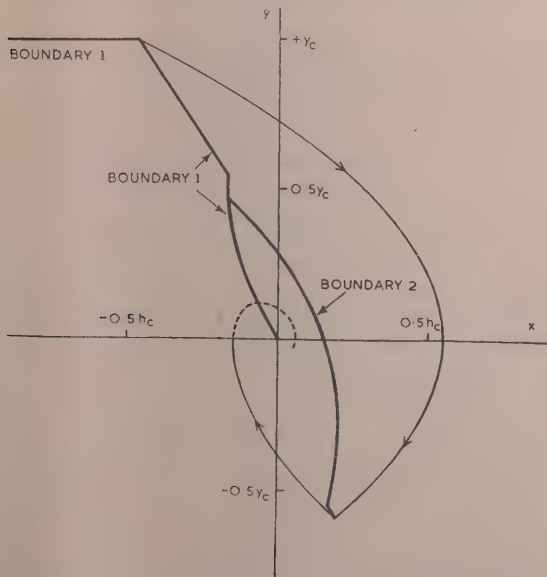


Fig. 10.—Boundaries with velocity limiting to $y_{max} = y_c$, showing the final section of the trajectory for large input steps.

$$\beta = 3; V = 0.3V_{max} = 0.2h_c.$$

alone, except that for large input steps the first boundary is now the horizontal line $y = y_{max}$. Thus extra stabilization is obtained for large input steps for which the final response becomes independent of input amplitude, since the output runs in along the horizontal portion of boundary 1 (Fig. 10). The control signal is given by

$$-x + I = \tau y + v_c$$

but during velocity limiting the signal $-x + I$ remains constant at a limiting value $-K\tau y_c$. Viscous and coulomb friction being neglected, acceleration occurs until $v_c = 0$, after which constant velocity ensues. Thus, during velocity limiting, $\tau y_{max} = K\tau y_c$ and

$$y_{max} = Ky_c (= \text{constant})$$

The time response in this state is given by $x = y_{max}t$, from which

$$\begin{aligned} x &= Ky_c t = K \frac{2\omega_0 h_c t}{\alpha \beta} \\ &= 2Kh_c \frac{\omega_0 t}{\alpha \beta} = 2Kh_c T_n \end{aligned}$$

as will be seen from Section 9.1

Time responses for a particular system with parameters corresponding to those of Fig. 10, and with $K = 1$, are shown in Fig. 11.

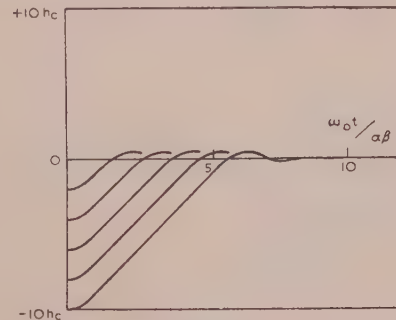


Fig. 11.—Responses with both velocity and integrator limiting.

$$y_{max} = y_c; \beta = 3; V = 0.3V_{max} = 0.2h_c.$$

(4) CLOSED-LOOP STABILITY BY THE DESCRIBING-FUNCTION TECHNIQUE

The use of the describing-function technique applied to the open loop is limited, as a quantitative analysis, to systems in which the closed loop may be opened so that the amplitude-dependent, frequency-independent non-linearity $N(h)$ is in series with the rest of the loop $F(\omega)$, which is dependent on frequency alone. The closed-loop gain is

$$m = \frac{N(h)F(\omega)}{1 + F(\omega)N(h)} = \frac{F(\omega)}{\frac{1}{N(h)} + F(\omega)}$$

Thus, provided $1/[N(h)] + F(\omega)$ cannot become zero, giving an infinite closed-loop gain, the system will remain stable for all input magnitudes. The Nyquist diagram of $-1/N(h)$ and $F(\omega)$ will give the phase and gain margins for any particular magnitude of input to the non-linear component, and will indicate instability by intersection of the two curves. The inclusion of a second non-linear component within $F(\omega)$ destroys its amplitude independence, and in general the describing-function analysis may only be carried out by considering the effect of one non-linearity at a time. This provides a qualitative analysis only, for the effect of two or more non-linear components in a servo system,

and there is the possibility that there may be some interaction not indicated by consideration of each one separately. In certain cases it is possible to make approximations enabling $F(\omega)$ to be constructed on the Nyquist diagram, while taking account of the non-linearity within $F(\omega)$.

(4.1) Torque Saturation and Integrator Limiting

The opened loop of Fig. 12A has the transfer function

$$F_1(\omega) = -\left(\frac{\omega_0}{\omega}\right)^2 \left\{ 1 + j \left[\alpha \left(\frac{\omega}{\omega_0} \right) - \frac{3}{\alpha \beta^2} \left(\frac{\omega_0}{\omega} \right) \right] \right\}$$

shown plotted on the Nyquist diagram in Fig. 12B, together with $-1/[N_1(h)]$ for the torque-limiting non-linearity.¹⁰ An

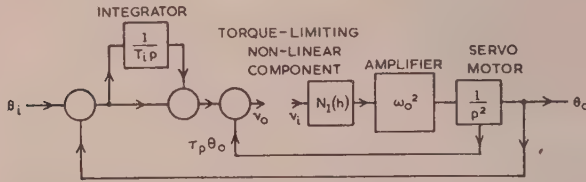


Fig. 12A.—Open-loop diagram for discussion of torque limiting.

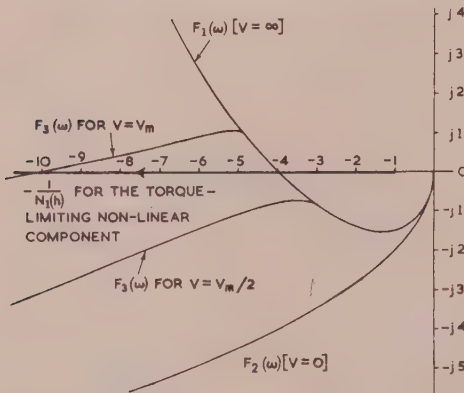


Fig. 12B.—Nyquist diagram showing the effect of torque and integrator limiting.

$$\alpha = 2; \beta = \sqrt{3}.$$

increase in $1/[N_1(h)]$ corresponding to increased signal v_i has the effect of reducing gain and phase margins until, with $1/[N_1(h)] = 4$, instability will occur. The frequency of oscillation will be given by

$$\alpha \left(\frac{\omega}{\omega_0} \right) = \frac{3}{\alpha \beta^2} \left(\frac{\omega_0}{\omega} \right), \text{ i.e. } \omega = \omega_0 \frac{\sqrt{3}}{\alpha \beta}$$

For comparison, the transfer function for the second-order system open loop (corresponding to $V = 0$) is shown, when

$$F_2(\omega) = -\left(\frac{\omega_0}{\omega}\right) \left(\frac{\omega_0}{\omega} + j\alpha \right)$$

For practical values of V the effective transfer function for the open loop will lie between the two curves, which are the extreme cases. By assuming that the signal from the non-linearity (Fig. 12A) has always its maximum value, independent of frequency, the effective transfer function including the limiting integrator may be obtained. This analysis will then hold provided that $1/[N_1(h)]$ is fairly large. Taking the fundamental of the maximum output to be $(2\sqrt{2}/\pi)v_m$ (r.m.s.), the amplitude

error is only 3% at $1/[N_1(h)] = 2$, decreasing for larger values of $1/[N_1(h)]$. The signal appearing at the input to the limiting integrator will now be $(2\sqrt{2}/\pi)v_m(\omega_0/\omega)^2$ in magnitude, at frequency ω . By reference to Fig. 14A the limiting-integrator describing-function for this input and frequency may be found. Let this value be the operator $(a - jb)$. Then

$$F_3(\omega) = -\left[\frac{\omega_0(\omega_0 + j\alpha)}{\omega} + \left(\frac{\omega_0}{\omega} \right)^2 (a - jb) \right]$$

$F_3(\omega)$ has been constructed by this method for two values of V , shown plotted on the Nyquist diagram in Fig. 12B. These curves coincide with the transfer function for the third-order system until limiting occurs in the integrator, when a tendency to approach the second-order system ($V = 0$) is apparent. For large input signals the system without limits on the integrator will become unstable for $1/[N_1(h)] > 4$ (Fig. 12B). With $V = 0$, and with $V = V_m/2$, stability is obtained for all input amplitudes. When $V = V_m$ conditional instability may occur, and continuous oscillation take place at the amplitude for which $1/[N_1(h)] = 10$.

(4.2) Frequency Analysis of Integrator Limiting

Fourier analysis of the limited-integrator output waveform shown in Fig. 13 yields for the peak value of the fundamental of the output:

$$I_0 = \frac{2}{\pi} \left\{ \left[(2V - B) \cos A + \frac{B}{4} (\cos 2A + 3) \right] - j \left[(2V - B) \sin A + \frac{B}{2} \left(\frac{\sin 2A}{2} + A \right) \right] \right\}$$

where $B = v/\omega CR$ and $A = \arccos [1 - (2V/B)]$, in the first and second quadrants. The limited-integrator describing-

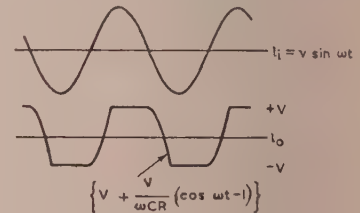


Fig. 13.—Limited integrator waveforms.

function calculated from the above expression is shown in Fig. 14A. This describing-function has also been obtained for an experimental limited integrator, and is shown in Fig. 14B plotted as curves of phase shift and modulus of describing-function against input amplitude. For a sinusoidal input whose amplitude increases in proportion to frequency so that without limits the output would be constant, the waveform appearing at the limited integrator output will be unaltered, except in frequency, provided the non-linear feedback provides a perfect short-circuit above the levels $\pm V$, and an open-circuit between these levels. The limited-integrator describing-function may therefore be used at any frequency if the scales of input and describing-function are calibrated in terms of ωCR and $1/\omega CR$ respectively, as in Figs. 14A and 14B. By including a differentiator before this limiting integrator the combination of the two units appears as a non-linearity dependent on amplitude alone. The polar curve of $-1/[N_2(h)]$ for this non-linearity is shown in Fig. 15a. In the open-loop diagram of Fig. 15A an extra integrator is

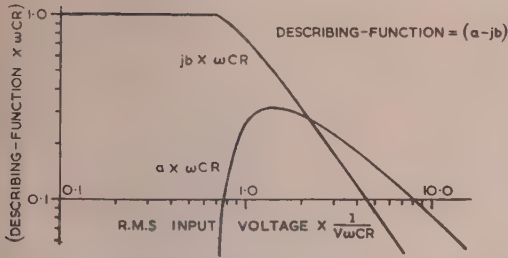


Fig. 14A.—Theoretical describing-function of the limiting integrator.

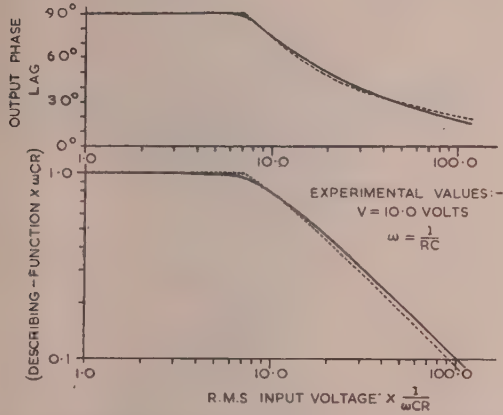


Fig. 14B.—Experimental describing-function of the limiting integrator. Theoretical curves are shown in broken lines.

included to neutralize the differentiator mentioned above. The frequency-dependent transfer-function from I_1 to I_0 is now

$$F_4(\omega) = \frac{-3}{\alpha\beta^2} \left(\frac{\omega_0}{\omega} \right) \frac{1}{\left\{ \alpha \left(\frac{\omega}{\omega_0} \right) - j \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right] \right\}}$$

and is shown on the diagram in Fig. 15B.

The stabilizing effect of the limiting integrator is apparent from the increase in gain and phase margins for large signals to the integrator. This does not take into account the effect of torque saturation. The occurrence of torque limiting (simultaneous with integrator limiting) has an unstabilizing effect, since the open-loop transfer function alters in shape when torque limiting occurs. $F_4(\omega)$ crosses the real axis at $\omega = \omega_0$ when $F_4(\omega) = -3/\alpha^2\beta^2 = -1/T_i^2\omega_0^2$.

Reduction of the effective value of ω_0^2 by torque limiting will increase the value of $F_4(\omega)$ at the real axis, reducing the gain and phase margins until instability occurs for $\omega_0^2 = 1/T_i^2$, if limiting in the integrator has not occurred. Stability for large signals is therefore dependent upon whether the stabilizing or unstabilizing influence is the more pronounced. Fig. 15C shows experimental curves of gain margin for two values of V , plotted against frequency, using the circuit and parameters of Figs. 15A and 15B respectively. For the particular circuit constants it is found that the onset of integrator limiting provides an immediate stabilizing action which more than offsets the unstabilizing action of torque limiting.

The effect of limits either at the input to or output from the error integrator is shown in Fig. 15B where the increase in $1/[N_3(h)]$ for this non-linearity does not give the same increase in phase margin as does the limited integrator in Fig. 6A. Owing

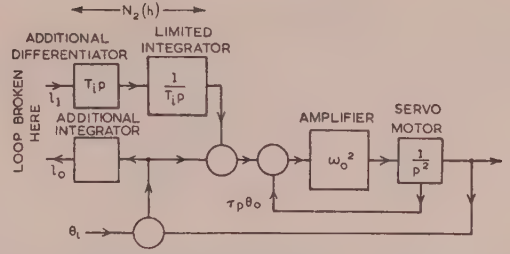


Fig. 15A.—Open-loop diagram for discussion of effect of integrator limiting.

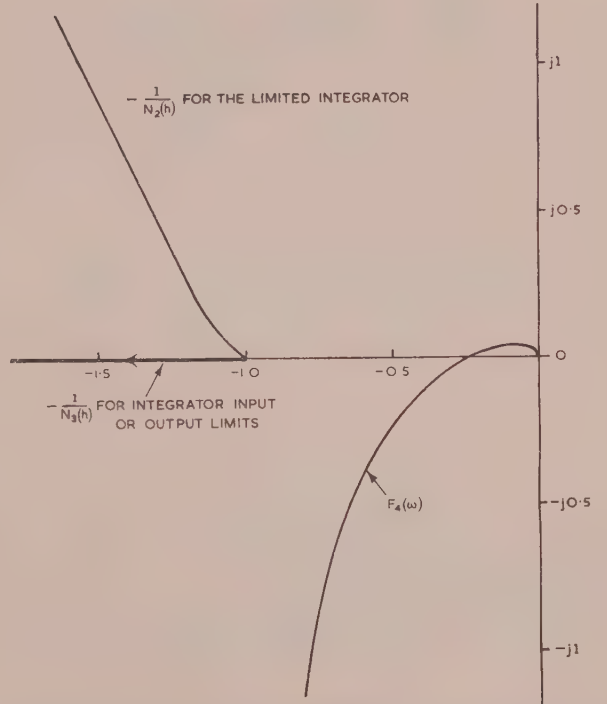


Fig. 15B.—Nyquist diagram showing effect of integrator limiting.

$$\alpha = 2; \beta = \sqrt{3}.$$

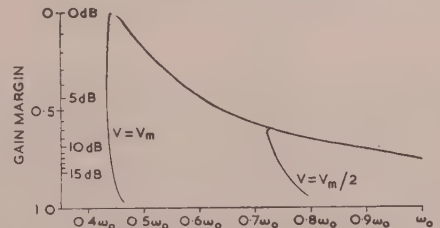


Fig. 15C.—Experimental curves to show the variation of gain margin with frequency.

$$\alpha = 2; \beta = \sqrt{3}.$$

to the reduction of phase shift for large signals to the limited integrator, conditional instability in a system is less likely to occur when using this integrator than when using input or output limits.

(4.3) Velocity Limiting

The open-loop diagram of a system employing velocity limiting and error integration is shown in Fig. 16A, where

$$F_5(\omega) = \frac{-\left(\frac{\omega_0}{\omega}\right)^2 \left[3/\alpha\beta^2 + j\left(\frac{\omega}{\omega_0}\right) \right]}{\left[\alpha + j\left(\frac{\omega}{\omega_0}\right) \right]}$$

The Nyquist diagram of $F_5(\omega)$ is given in Fig. 16B, together with the values of $-1/[N_4(h)]$ for the velocity-limiting non-

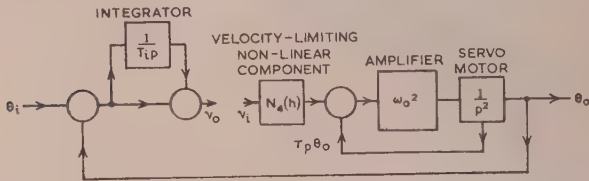


Fig. 16A.—Open-loop diagram for discussion of effect of velocity limiting.

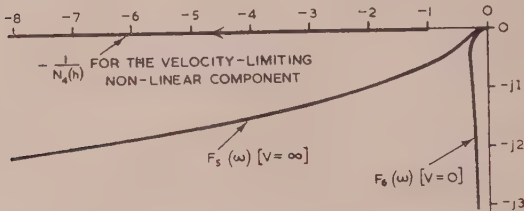


Fig. 16B.—Nyquist diagram showing effect of velocity limiting.
 $\alpha = 2$; $\beta = \sqrt{3}$.

linearity. The effect of increased signals is a reduction of the phase margin, giving reduced stability. For the second-order system, corresponding to $V = 0$,

$$F_6(\omega) = -\left(\frac{\omega_0}{\omega}\right) \frac{1}{\left(\frac{\omega}{\omega_0} - j\alpha\right)}$$

shown on the Nyquist diagram in Fig. 16. An increase in the input signal will now give an increase in the phase margin and increased stability. The actual curve of $F(\omega)$ for the system when the integrator has limits, will lie somewhere between these two curves if the effect of torque limiting is neglected.

(5) STOCHASTIC INPUT TEST

The response of the system to a noise input of given power and defined bandwidth has been obtained experimentally for several values of V , and is shown in Fig. 17 as an error-spectrum response. Such a test corresponds more nearly to a general input than does a step or frequency response, and is included for this reason.

The describing-function for the torque-saturation non-linearity for a small sinusoidal signal in the presence of a large noise signal is found to be independent of frequency,¹¹ because cross-modulation products can be neglected.¹² Thus, in accordance with the assumptions of the describing-function technique applied to noise signals, the error-spectrum responses for $V = 0$ and for V very large are found to be the same as those obtained for simple sinusoidal input to the linear system but, with reduced amplifier gain. A particular case of the reduction in gain of the saturation characteristic is given in the experimental

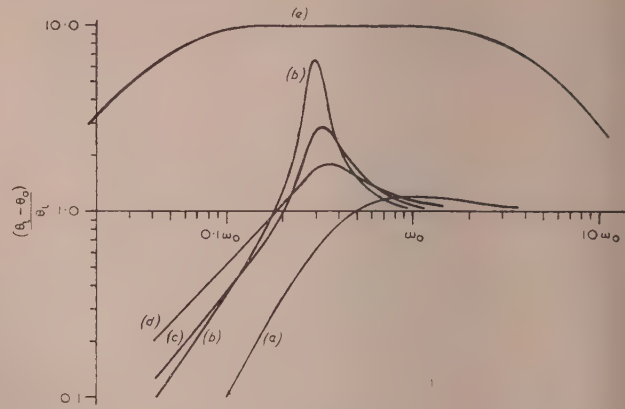


Fig. 17.—Experimental spectral responses to a noise-signal input.

$\alpha = 2$; $\beta = 3$. Noise input voltage = $0.5h_c$ r.m.s.

(a) Linear system response.

(b) $V = 0.6V_{max} = 0.4h_c$.

(c) $V = 0.3V_{max} = 0.2h_c$.

(d) $V = 0$.

(e) Input spectrum.

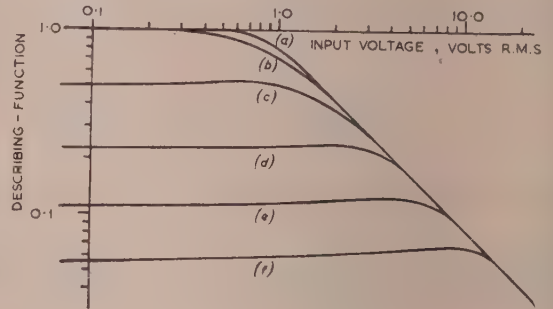


Fig. 18.—Experimental curves to show modification of torque-limiting non-linearity describing-function by an added sinusoidal signal of high frequency. $v_m = \pm 1.0$ volt.

- (a) Added signal 0 volt.
- (b) Added signal 0.5 volt r.m.s.
- (c) Added signal 1.0 volt r.m.s.
- (d) Added signal 2.0 volts r.m.s.
- (e) Added signal 4.0 volts r.m.s.
- (f) Added signal 8.0 volts r.m.s.

curves in Fig. 18, in which the describing-function is shown for several values of an added sinusoidal signal of much higher frequency. The reduction in amplifier gain due to the noise signal at the input to the saturation non-linear component may cause instability if the integrator is not limited. This would be so in the responses of Fig. 17. For a noise input which would only just cause instability the frequency of continuous oscillation is $\omega = \omega_0\sqrt{3/\alpha\beta}$ (Section 4.1), which for the particular system occurs at $\omega = \omega_0(1/2\sqrt{3})$. Peaks in the frequency-spectrum error response may therefore be expected to occur at about this frequency.

The system will obviously become more stable for a given amplitude of noise input if the higher frequencies do not exist in the input, since it is these input frequencies, occurring as error signals since the system cannot respond to them, which reduce the effective gain of the amplifier, and therefore the stability.

Experimental curves are given in Fig. 19 to show the behaviour of the limiting integrator with a noise input corresponding approximately in magnitude and frequency spectrum to the error signals in Fig. 17. The reduction in integrator phase shift and describing-function at the lower frequencies provides the necessary stabilizing influence with large noise-signal inputs.

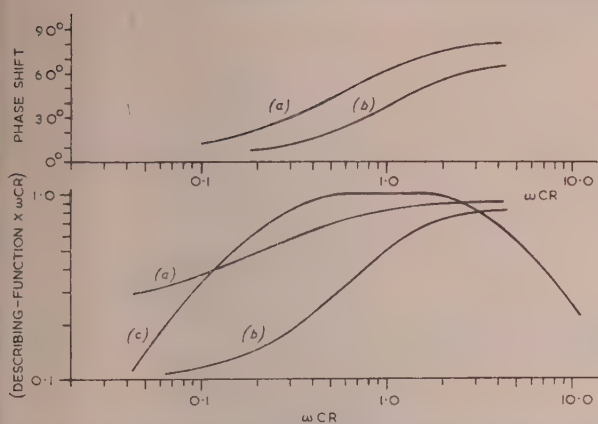


Fig. 19.—Experimental curves to show the limited-integrator describing-function for a noise input signal of given bandwidth.

- (a) Noise input = $3\sqrt{3}\cdot6$ volts r.m.s.
 (b) Noise input = $8\sqrt{3}\cdot6$ volts r.m.s.
 (c) Input spectrum.

Experimental values: $V = 3\cdot6$ volts; $RC = 0\cdot0025$ sec.

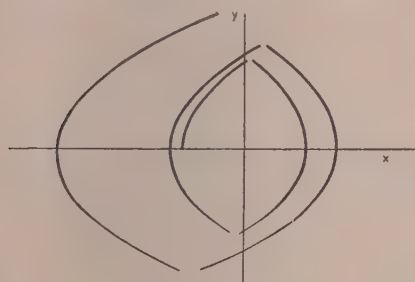


Fig. 20A.—Experimental phase-plane response for a step input just greater than h_c . $\alpha = 2$; $\beta = 3$.

The linear regimes have been automatically blanked out.

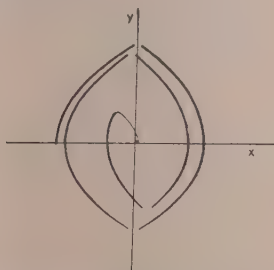


Fig. 20B.—Experimental phase-plane response for a step input just less than h_c . $\alpha = 2$; $\beta = 3$.

The linear regimes have been automatically blanked out.

(6) RESULTS

Figs. 8A, 8B, 8C, 9A, 9B, 9C and 11, in which the effect of passage through the linear regime is neglected, enable the response of a particular system to be determined, and give the necessary information required to stabilize against large input steps. By the addition of non-linear components many of the adverse effects associated with error integration in a torque-limited system are eliminated.

The phase-plane boundaries and step responses for the systems have been checked by simulator. Figs. 20A and 20B show responses in which the linear regime has been automatically blanked out, for input steps just greater than and just less than h_c .

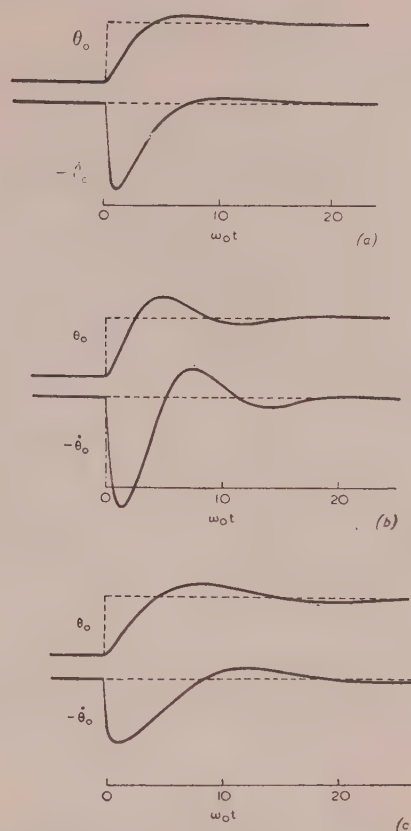


Fig. 21.—Linear response to a step input.

- (a) $\alpha = 2$; $\beta = 3$.
 (b) $\alpha = 2$; $\beta = \sqrt{3}$.
 (c) $\alpha = 2\sqrt{3}$; $\beta = \sqrt{3}$.

Figs. 21(a), 21(b) and 21(c) are included for reference, being linear responses for the particular values of α and β considered.

All the transient response behaviour has been further checked on a small servo model.

(7) CONCLUSIONS

The main part of the paper has shown how the transient behaviour of an integral controlled system with torque saturation may be obtained by means of phase-plane methods. This is extended to include additional non-linear elements to prevent high-amplitude, low-frequency oscillation and to limit maximum possible velocities.

The stability of these systems has been investigated by the describing-function technique and by frequency-response analysis. These show qualitatively the kind of result to be expected from the phase-plane analysis.

Experimental results for a noise input show that stability can always be maintained by reducing the integrator limits by an amount dependent on both bandwidth and magnitude of the noise signal.

(8) REFERENCES

- (1) PORTER, A.: 'An Introduction to Servomechanisms' (Methuen, London, 1950).
- (2) WEST, J. C.: 'A System utilizing Coarse and Fine Position-Measuring Elements simultaneously in Remote-Position-Control Servo Mechanisms', *Proceedings I.E.E.*, Paper No. 1185 M, September, 1951 (99, Part II, p. 135).

- (3) ROUTH, E. J.: 1859. 'A Treatise on the Dynamics of Rigid Bodies' (Macmillan, 6th ed., 1905).
- (4) WEST, J. C.: 'Servomechanisms' (English Universities Press, London, 1953).
- (5) LIENARD, A.: 'Étude des oscillations entretenues', *Revue Générale de l'Électricité*, 1928, **23**, p. 901.
- (6) BROWN, G. S., and CAMPBELL, D. P.: 'Principles of Servomechanisms' (Chapman and Hall, 1948, p. 92).
- (7) WEST, J. C., DOUCE, J. L., and NAYLOR, R.: 'The Effects of the Addition of some Non-Linear Elements on the Transient Performance of a Simple R.P.C. System possessing Torque Limitation', *Proceedings I.E.E.*, Paper No. 1549 M, August, 1953 (**101**, Part II, p. 156).
- (8) WEST, J. C., and DALTON, I. R.: 'The Step-Function Response of an R.P.C. Servo possessing Torque Limitation', *ibid.*, Paper No. 1576 M, November, 1953 (**101**, Part II, p. 166).
- (9) JOHNSON, E. C.: 'Sinusoidal Analysis of Feedback-Control Systems containing Non-Linear Elements', *Transactions of the American I.E.E.*, 1952, **71**, Part II, p. 169.
- (10) GRIEF, H. D.: 'Describing Function Method of Servomechanism Analysis applied to most commonly encountered Non-Linearities', *ibid.*, 1953, **72**, Part II, p. 243.
- (11) WEST, J. C., and NIKIFORUK, P.: 'The Behaviour of Non-Linear Systems subjected to Random Inputs' (to be published).
- (12) WEST, J. C., and NIKIFORUK, P.: 'The Response of R.P.C. Systems with Hard-Spring Non-Linear Characteristics to Step-Function and Random Inputs', *Proceedings I.E.E.*, Paper No. 1728 M, October, 1954 (**102** B, p. 575).

(9) APPENDICES

(9.1) The Phase-plane Boundaries

Under conditions of saturation the motion of the output is given by

$$\frac{d^2\theta_0}{dt^2} = y \frac{dy}{dx} = \pm a$$

where $a = \omega_0^2 v_m$.

Thus by direct integration (see Section 2.3)

$$y^2 - y_0^2 = \pm 2a(x - x_0)$$

where x_0 and y_0 are the initial values of position and velocity.

These are parabolae in the xy plane and continue provided that the servo motor remains saturated.

The boundary between linear and saturated conditions is given by

$$v_c = \pm v_m$$

$$\text{where } v_c = (\theta_i - \theta_0) - \tau \frac{d\theta_0}{dt} + I$$

Consider the response to a step of magnitude h from $\theta_i = \theta_0 = -h$ to the origin with the system initially at rest. If h/v_m is greater than unity, saturation will ensue immediately.

The value of v_c is determined by consideration of the three components in turn:

(a) *The error $\theta_i - \theta_0$.*

This becomes $\theta_i - \theta_0 = -x$, since $\theta_i = 0$ for $t > 0$.

(b) *Velocity damping term.*

$$\frac{\tau d\theta_0}{dt} = \tau y$$

(c) *Integrator output.*

$$\begin{aligned} I &= \frac{1}{T_i} \int_0^t (\theta_i - \theta_0) dt \\ &= -\frac{1}{T_i} \int_0^t x dt \\ &= -\frac{1}{T_i} \int_0^t x \frac{dt}{dy} dy \\ &= -\frac{1}{T_i} \int_0^y \frac{x}{a} dy \quad \text{since } \frac{dt}{dy} = \frac{1}{a} \end{aligned}$$

Integrating by parts, we have

$$I = -\frac{xy}{T_i a} + \frac{1}{T_i} \int_0^y \frac{y}{a} dy$$

But $y dy = a dx$, from which

$$\begin{aligned} I &= -\frac{xy}{T_i a} + \frac{1}{T_i} \int_0^y \frac{y^2}{a^2} dy \\ &= -\frac{xy}{T_i a} + \frac{y^3}{3a^2 T_i} \quad \dots \dots \dots (15) \end{aligned}$$

Hence, by addition of these three terms, the equation of the boundary in the xy plane is obtained:

$$v_m = -x - \tau y - \frac{xy}{T_i a} + \frac{y^3}{3a^2 T_i} \quad \dots \dots \dots (16)$$

If the actual width of the linear regime v_m is assumed to be small compared with the magnitudes under consideration, the discussion is simplified by making v_m zero. Thus

$$y^3 - 3axy - 3a^2 T_i \tau y - 3a^2 T_i x = 0$$

is the equation of the boundary for which change-over of saturation from $+a$ to $-a$ occurs.

This cuts the y axis (i.e. at $x = 0$) at

$$y_c^2 = 3a^2 \tau T_i$$

The input step $-h_c$ which would give a trajectory to cut the boundary at this point is given by

$$y_c^2 = 2a(x + h_c)$$

and for $x = 0$

$$h_c = \frac{3}{2} a \tau T_i$$

The boundary equations can be unified for all systems by a change of the variable.

Let $H = x/h_c = x/(3/2 a \tau T_i)$

and $u = y/y_c = y/a\sqrt{(3\tau T_i)}$

Then the boundary equations become

$$H = \frac{2(u^2 - 1)u}{3au + \beta} \quad \dots \dots \dots (17)$$

where $\beta = \sqrt{(3\tau T_i)}$

It is seen that the boundaries all pass through the points (0, 0) and (0, 1) in the Hu plane and are independent of α .

The time solution for the torque-saturated regions is given from eqn. (8):

$$\begin{aligned} x &= -h + \frac{1}{2} a t^2 \\ x + h &= \frac{h_c t^2}{3\tau T_i} \end{aligned}$$

Now

$$\alpha^2 \beta^2 = \tau^2 \omega_0^2 \frac{3T_i}{\tau} = 3\tau T_i \omega_0^2$$

Hence

$$\frac{x+h}{h_c} = \left(\frac{\omega_0 t}{\alpha\beta}\right)^2$$

and a time scale $T_n = \frac{\omega_0 t}{\alpha\beta}$ can be used.

The time to reach the critical velocity y_c for an initial step h_c occurs at $x = 0$, i.e.

$$\left(\frac{\omega_0 t}{\alpha\beta}\right)^2 = T_n^2 = \frac{h_c}{h_c} = 1$$

(9.2) Graphical Determination of Boundaries

Under saturation conditions

$$\frac{d^2\theta_0}{dt^2} = \frac{dy}{dt} = a \quad (-a \text{ may, of course, be the alternative form})$$

$$y \frac{dy}{dx} = a$$

and

$$y^2 - y_0^2 = 2a(x - x_0)$$

Let

$$I = I_0 - \frac{1}{T_i} \int_0^t \theta_0 dt, \text{ assuming } \theta_i = 0$$

$$= I_0 - \frac{1}{T_i} \int_0^t x dt$$

$$= I_0 - \frac{1}{T_i} \int_0^t \frac{x}{a} dy$$

Integrating by parts,

$$I = I_0 - \frac{1}{T_i} \left[\frac{xy}{a} \right]_0^t + \frac{1}{T_i} \int_0^t \frac{y dx}{a}$$

But $y dy = a dx$; therefore

$$\begin{aligned} I &= I_0 - \frac{1}{T_i} \left[\frac{xy}{a} \right]_0^t + \frac{1}{T_i} \int_0^t \frac{y^3}{3a^2} dy \\ &= I_0 - \frac{xy}{aT_i} + \frac{x_0 y_0}{aT_i} + \frac{y^3}{3a^2 T_i} - \frac{y_0^3}{3a^2 T_i} \end{aligned} \quad (18)$$

and this is the integrator output.

A boundary is reached each time the control signal v_c becomes zero, i.e.

$$v_c = -x - \tau y + I$$

The equation enables I to be plotted on the xy phase plane.

Initially, for a step function h from $x_0 = -h$, the conditions are

$$I_0 = 0$$

$$y_0 = 0$$

and hence

$$I = -\frac{xy}{aT_i} + \frac{y^3}{3a^2 T_i}$$

This is shown in Fig. 4, where I is plotted vertically against x for each point x, y of the trajectory

$$y^2 = 2a(x + h)$$

The curve of τy against x is also plotted.

Thus, at the point C_1 the integrator output is $B_1 C_1$, the velocity signal $\tau \theta_0$ is $A_1 C_1$, and if by inspection these points are chosen so that

$$A_1 B_1 = A_1 D_1$$

we then have

$$I - \tau y - x = 0$$

and this is a point on the boundary.

New curves must now be constructed from B_1 and A_1 with $-a$ as the acceleration and initial conditions $x_0 y_0 I_0$ that apply at $A_1 B_1$ and C_1 .

Thus the second boundary would be given by

$$A_2 D_2 = A_2 B_2$$

This method can be continued until all the change-over points for a given trajectory have been obtained.

(9.3) Graphical Determination of the Phase-Plane Trajectory for the Linear System

The slope of the trajectory in the xy plane is a single-valued function of x and y for second-order systems. For a third-order system this uniqueness is lost; the slope at a particular point depends also on the acceleration. There is still, however, only one trajectory for fixed initial conditions of position, velocity and acceleration. In many circumstances the initial acceleration is known or can be found by the initial-value theorem.⁶

For the particular case of the system with integral control, eqn. (4) can be written

$$\left[\frac{p^2}{\omega_0^2} + (\tau + \tau_v)p + 1 + \frac{1}{T_i p} \right] \theta_0 = \left(1 + \frac{1}{T_i p} \right) \theta_i$$

Let

$$(\theta_0 - \theta_i) = x$$

$$\frac{d\theta_0}{dt} = y$$

$$\frac{d^2\theta_0}{dt^2} = \frac{dy}{dt} = \frac{dy}{dx} \frac{dx}{dt} = \frac{dy}{dx} (y - \dot{\theta}_i) = y \frac{dy}{dx}$$

if $\dot{\theta}_i$ is zero as in an applied step for $t > 0$.

$$\text{Hence } \frac{y}{\omega_0^2} \frac{dy}{dx} + (\tau + \tau_v)y + x + \frac{1}{T_i} \int x dt = I_0$$

where I_0 is some constant of integration dependent on the initial conditions. If these, at time $t = 0$, are

$$x = x_0$$

$$y = y_0$$

and

$$y \frac{dy}{dx} = z_0 \text{ (the acceleration)}$$

then

$$I_0 = x_0 + (\tau + \tau_v)y_0 + \frac{z_0}{\omega_0^2}$$

and is known.

Let

$$I = -\frac{1}{T_i} \int x dt + I_0$$

Then

$$\frac{dI}{dt} = \frac{dI}{dx} \frac{dx}{dt} = y \frac{dI}{dx} = -\frac{x}{T_i}$$

and

$$\frac{dI}{dx} = -\frac{x}{yT_i}$$

The slope of the trajectory is given by

$$\frac{dy}{dx} = -\frac{\omega_0^2 x + (\tau + \tau_v)y - I}{y}$$

Using small straight-line-segment approximation it is possible to plot two trajectories, one for the Ix plane and the other for the xy plane. Starting from the known initial values, the first value of I may be obtained and used to find the isocline dy/dx and hence the next point (x, y) of the trajectory. This in turn is used to obtain the isocline dI/dx and then the next value of I .

This technique is very useful in obtaining the trajectory of the non-linear system during its passage through the linear regime because the conditions of x, y, z are known on the boundary.

THE USE OF ETHYLENE DIAMINE TARTRATE FOR PIEZO-ELECTRIC FILTER ELEMENTS

By J. BIRCH, Graduate, A. G. FRITH, Associate Member, A. C. L. FERGUSON, Graduate,
R. H. A. MILES, M.A., B.Sc., A.Inst.P., and J. F. WERNER, B.Sc.

(The paper was first received 17th January, and in revised form 8th March, 1956. It was published as an INSTITUTION MONOGRAPH in June, 1956.)

SUMMARY

Crystal resonators suitable for use in telephone system channel filters have been made from ethylene diamine tartrate (e.d.t.). Methods used in the fabrication and mounting of these units are described. The circuits used in the determination of the electrical parameters are given, and the results of a large number of measurements are outlined. Freedom from unwanted resonances over the desired frequency bands combined with the required characteristics has been achieved by suitably dimensioning the plates, and by applying electrodes to a part of the crystal surface only. Increased activity for crystals mounted in air can be obtained by the use of suitably placed reflectors.

LIST OF PRINCIPAL SYMBOLS

- f_a = Parallel resonant frequency of crystal.
 f_r = Series resonant frequency of crystal.
 N = Frequency constant of crystal.
 A = Curvature constant of frequency/temperature characteristic.
 T_0 = Inversion temperature of frequency/temperature characteristic.
 C_m = Equivalent series capacitance of crystal.
 L_m = Equivalent series inductance of crystal.
 R_m = Equivalent series resistance of crystal.
 C_0 = Shunt capacitance of crystal.
 K = Inductance constant of crystal.
 α = Temperature-coefficient of crystal inductance.
 l = Length of crystal plate.
 w = Width of crystal plate.
 t = Thickness of crystal plate.
 l' = Length of plating.

(1) INTRODUCTION

Multi-channel carrier telephone systems incorporate large numbers of electric wave filters which must be constructed to very precise specifications. These filters must include sharply tuned resonating circuits having high stability of resonant frequency, inductance and capacitance, over the range of temperature met with in operation. Such a resonating circuit may be simulated by a piezo-electric crystal element which is cut at a precise orientation from the parent crystal, dimensioned, and so mounted as to be free to vibrate at its fundamental frequency. Excitation is achieved by thin metal electrodes attached to the surfaces of the element.

For some time past, quartz has been used with success for making such resonator crystals. On account of the comparatively low frequencies involved, the elements are correspondingly large, and, since each must be free of all flaws or twinning, they require material of high quality. Natural quartz for use in piezo-electric

applications must be imported, so there are both economic and strategic reasons for seeking an acceptable alternative material.

Ethylene diamine tartrate (abbreviated herein to e.d.t.) is one of the more promising substitutes for quartz so far investigated. The discovery of its suitability and the preliminary investigation into methods of growing and utilizing it were made by the Bell Telephone Laboratories.¹⁻⁷ It may be grown in large crystals having a good yield of flawless material, and the methods adopted for the growth of crystals used in the experiments which are the principal subject of the paper have been described elsewhere by L. A. Dauncey.⁸⁻¹⁰ As in the case of quartz, it is possible to obtain resonators having low or zero temperature-coefficients of frequency by suitable orientation with respect to the parent crystal. The derivation of these cuts has been carried out mainly by Mason,¹¹ Jaffe,¹² Bond¹³ and Bechmann.¹⁴

As compared with quartz, however, e.d.t. offers several practical difficulties. It is water-soluble, and, although not deliquescent, is readily harmed by moisture. Because of its pronounced anisotropic thermal expansion, coupled with its fragility, large crystals are liable to crack if subjected even to small temperature gradients. Special care is therefore needed in handling and processing the material.

A further difficulty which may severely limit the use of e.d.t. crystals on a large scale is that, while the crystals themselves are harmless, certain hazards are connected with their preparation and processing which have been known to result in dermatitis.

Although suitable techniques have been devised for designing and fabricating low-frequency resonators from e.d.t., it has not been found suitable for high-frequency applications. This is due partly to the difficulty of making very thin plates and partly to the greater number of unwanted resonances produced as compared with quartz. This latter is not surprising in view of the lower crystal symmetry of e.d.t. and the consequent greater importance of cross-coupling effects, further accentuated by the relatively high piezo-electric activity.

The electrical parameters of a crystal unit depend on the dimensions, electrode form and method of mounting of the plate. Previously published work does not give sufficient information to enable a unit with a given performance to be produced. The present investigation was therefore undertaken so that crystals could be manufactured with predictable characteristics, including freedom from unwanted resonances over a given frequency range.

(2) CRYSTAL CUTS AND ORIENTATION

An idealized e.d.t. crystal is shown in Fig. 2(a), which also shows the positions of the three conventional crystallographic axes a , b , and c and the orthogonal axes X , Y , and Z . The principal zero temperature-coefficient cuts considered in the paper are shown in Figs. 2(b)–2(f). Using the standard convention¹⁵ the orientations are also given.

The 0° Y-cut, the B-cut and the A-cut all operate in their fundamental mode of longitudinal vibration. The 0° Y-cut is

Correspondence on Monographs is invited for consideration with a view to publication.

The paper is a communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

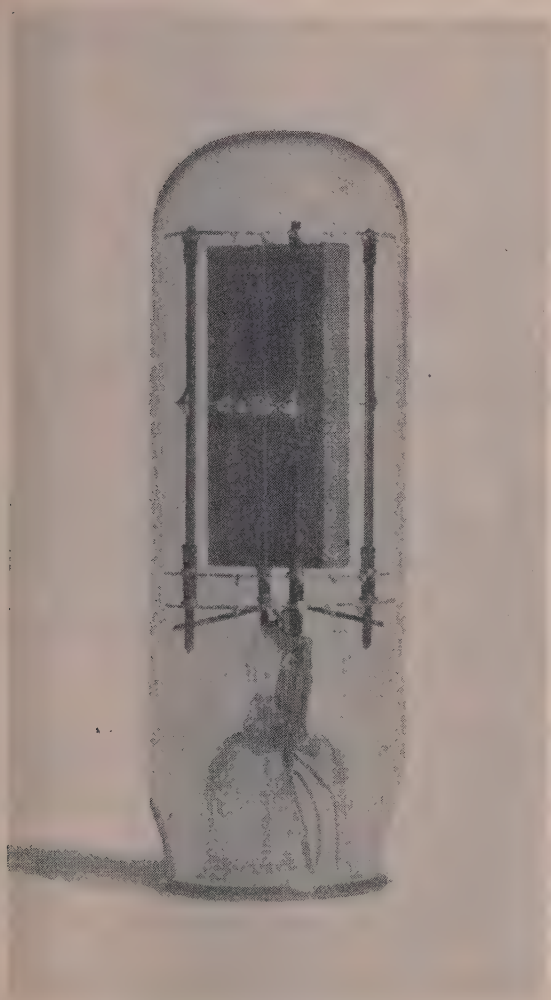


Fig. 1.—Complete crystal unit.

the most widely used and possesses suitable frequency/temperature characteristics. The B-cut shows similar characteristics and may be of use for plates having equivalent inductances which cannot be obtained with the Y-cut because of interference by unwanted resonances.

The A-cut has a more complex orientation for production-scale cutting, and shows unwanted resonances closer to the fundamental frequency. The A-, B-, and Y-cuts can all be used for filters in the frequency range 60–100 kc/s.

The Bechmann 77°-cut is employed as a face-shear plate, using centre-point mounting. It suffers, however, from the disadvantage that a cleavage plane almost coincides with the major surfaces, making wire mounting difficult, and little work was done with this cut in the present investigation.

Fig. 2(f) shows the orientation of a crystal suitable for use in its high-frequency shear mode.¹¹

(3) THE FABRICATION AND MOUNTING OF FILTER CRYSTALS

Fig. 1 shows a completed crystal unit mounted in an evacuated glass bulb. It will be seen that two pairs of electrodes are applied to the e.d.t. plate, this method being commonly used in the lattice type of filter circuit, where the two halves of the crystal

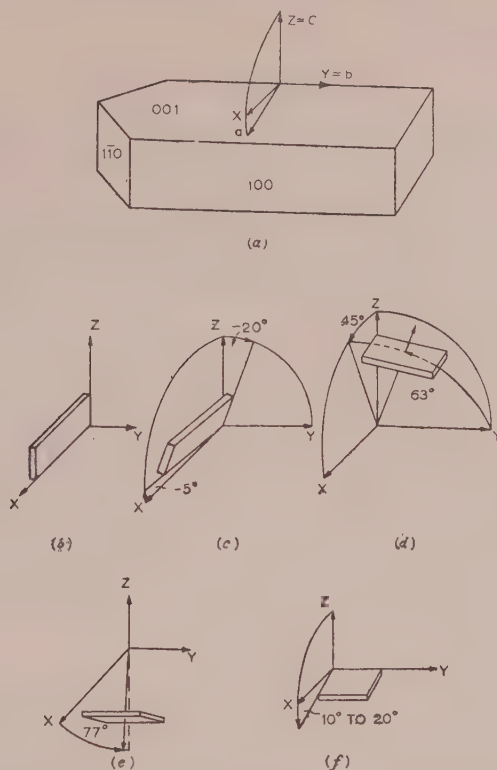


Fig. 2.—Orientation of zero temperature-coefficient cuts according to United States nomenclature.

- (a) Parent crystal.
- (b) Y-cut (YX).
- (c) B-cut (YXlt) $\pm 20^\circ/-5^\circ$.
- (d) A-cut (YZltw) $45^\circ/63^\circ$.
- (e) Bechmann 77°-cut (XYlt) 77° .
- (f) Thickness shear cut (ZYlt) 10° to 20° .

element appear in two arms of the bridge. The number of crystal elements needed to obtain the desired filter characteristic is thus halved.

(3.1) Processing

The methods used for cutting and lapping are basically similar to those employed for quartz crystals. Although it is possible to employ such techniques as wet-string cutting, it was found preferable to use diamond-loaded saws, thus reducing the amount of subsequent dimensioning required. Oil is used as a lubricant and coolant.

The final dimensioning of the crystal blanks is accomplished by grinding on stationary cast-iron laps, using 400-mesh Carborundum as the abrasive, with an oil vehicle. The major surfaces are dimensioned in a fixture which maintains the upper surface of the crystal parallel to the lap. The edges are dimensioned in a fixture which maintains the major surfaces perpendicular to the lap.

At each stage in the production of the blank its orientation is checked by measurement on an X-ray goniometer. E.D.T. gives several strong reflections that are suitable for this purpose. By means of these techniques, tolerances of ± 0.01 mm on linear dimensions and $\pm 5'$ on orientation may be readily achieved. These are, however, unnecessarily close for many applications.

An improvement in the Q-factor of e.d.t. crystals can be obtained by etching. For this process, mixtures of water and methylated spirit or water and ethylene glycol were employed, the former being preferred for economic reasons. In both cases

the etching is terminated by transferring the crystal to ethylene glycol as a stop bath.

(3.2) Wire Mounting and Plating

Crystal units for use in filters are normally wire-mounted. Quartz plates have the wires soldered to a layer of silver and lead borate which is fused on the plate at high temperature. E.D.T., however, decomposes at about 120°C in air and an adhesive which hardens below this temperature must therefore be used. Phosphor-bronze wires 0.008 in in diameter are used for crystal mounting. Metal heads 1 mm in diameter are soldered or welded to the wire and an adhesive is used to attach these heads to the crystal surface.

The thermal expansion of the crystal varies between -11 and $+90$ parts in 10^6 per deg C in the major faces of a Y-cut plate, according to the direction. The adhesive employed must therefore be sufficiently pliant to accommodate the strain in the joint due to the difference in expansion between the wire head and the crystal. After many experiments the best adhesive was found to be a mixture of a thermoplastic polyvinyl ester and a thermo-setting Bakelite resin. Unmodified thermoplastic cements cannot be used because if they flow at the curing temperature, i.e. not over 120°C, they also tend to soften at a high operating temperature (60°C). By these techniques crystals were fabricated which were adequately robust and capable of being operated at up to 60°C without mechanical deterioration or large changes in electrical parameters.

After attachment of the wire, silver electrodes are deposited on the crystal by an evaporation method. Sputtering cannot easily be used for electrode deposition because of the rapid temperature rise to which the crystal is subjected. Many of the experimental plates were mounted in evacuated glass bulbs, although it is possible to mount the crystals on mica cages in sealed containers filled with dry air.

In order to minimize the damping effect of the wire attachments, weights in the form of solder balls are fixed to the mounting wires at a distance of one-quarter or three-quarters of a wavelength along the wire.¹⁶

(4) MEASUREMENT OF CRYSTAL PARAMETERS

The equivalent electrical circuit of a piezo-electric crystal for frequencies near to that of the main response is shown in Fig. 3A.

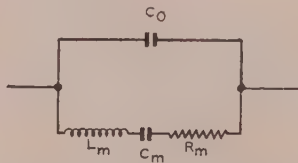


Fig. 3A.—Equivalent circuit of piezo-electric crystal.

The equivalent series inductance, L_m , may be considered to represent the inertia of the crystal, the equivalent series capacitance, C_m , to represent the compliance. The equivalent series resistance, R_m , represents losses due to the motion of the crystal. These include both internal loss in the crystal and external losses arising from the mounting and air damping. The capacitance C_0 may be taken as the total static capacitance across the crystal.

Such a circuit has an impedance characteristic as shown in Fig. 3B. At the frequency f_r , C_m resonates with L_m and the crystal behaves as a resistance R_m in parallel with C_0 . At a higher frequency, f_a , the system acts as a parallel-resonant circuit in which L_m resonates with the series combination of C_m and C_0 . The crystal has a high resistive impedance at this frequency, which is approximately equal to $1/(4\pi^2 f_a^2 C_0^2 R_m)$.

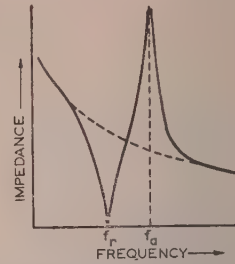


Fig. 3B.—Impedance/frequency characteristic.

The two frequencies f_r and f_a are known as the series-resonant and parallel-resonant frequencies of the crystal. It may be shown that, to a high degree of approximation,

$$f_a - f_r = f_r \frac{C_m}{2C_0}$$

The factors of most interest in designing a crystal resonator for use in a filter are the resonant frequency, the inductance and the series resistance, and the variation of these quantities with temperature.

For the important case of a crystal vibrating in its longitudinal mode, the frequency is related to the length l of the plate by the equation,

$$f_r l = N$$

where N is a constant known as the frequency constant. For crystals of finite width this is only true to a first degree of approximation, and it is shown in Section 5.1.2 that the value of N to be used depends on the width/length ratio of the resonator plate.

The inductance of the plate is related to the length l , width w and thickness t of the crystal to a first order of approximation by the equation

$$L_m = K \frac{lt}{w}$$

where K is a constant known as the inductance constant. Again, it has been shown experimentally that there is a slight variation of K with the width/length ratio.

(4.1) Measurement of Equivalent Inductance

Two methods were employed for measuring the equivalent series inductance L_m . Fig. 4 shows a schematic form of the

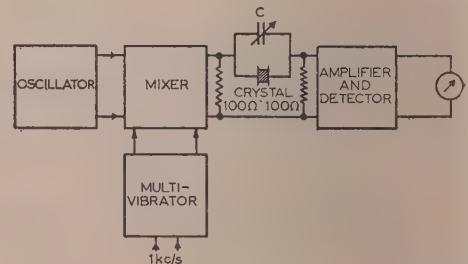


Fig. 4.—Block schematic of arrangement for determination of L_m .

circuit used in the first method. The crystal is connected in a π -network and is shunted by an accurately calibrated variable capacitor C which has a very low minimum capacitance (about $0.05 \mu\text{F}$).

The multivibrator may be set to deliver any one of a series of sub-harmonics of the standard 1000 c/s signal (500, 333·3, 250 c/s, etc.). By adjustment of the signal-generator frequency on the high side of resonance a setting is found at which the lower sideband delivered from the mixer stage coincides with the series-resonant frequency f_r of the crystal. This is shown by a sharp maximum indication of the detector. The signal generator is then oscillating at a frequency $f_r + 1000/n_1$, where n_1 is a small whole number. With the multivibrator switched off, C is adjusted for minimum output from the detector, at which setting the signal-generator frequency is the parallel-resonant frequency for the crystal and capacitor combination. If C_1 is the capacitor reading,

$$f_a - f_r = \frac{1000}{n_1} = f_r \frac{C_m}{2(C_0 + C_1)}$$

from which

$$C_0 + C_1 = \frac{f_r C_m}{2} \frac{n_1}{1000}$$

Special precautions are taken to ensure that the signal used for this setting is free from harmonics.

By repeating the process with another sub-harmonic of 1000 c/s, say, $1000/n_2$, the unknown shunt capacitance of the crystal is eliminated, giving the relation,

$$C_1 - C_2 = \frac{f_r C_m}{2} \frac{n_1 - n_2}{1000}$$

and from this, since

$$L_m = \frac{1}{4\pi^2 f_r^2 C_m}$$

it follows that

$$L_m = \frac{1}{8\pi^2 f_r (C_1 - C_2)} \frac{n_1 - n_2}{1000}$$

In the second method, an inductor L and a capacitor C are connected across the crystal terminals, as shown in Fig. 5A. The reactance characteristic of this circuit is shown in Fig. 5B.

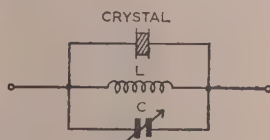


Fig. 5A.—Reactance arm for inductance measurement.

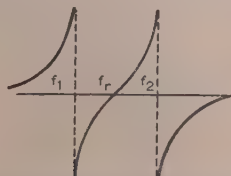


Fig. 5B.—Reactance/frequency characteristic.

For a given crystal and with the inductance L constant, f_1 and f_2 depend on the value of the capacitance C . The third point of zero reactance is the series-resonant frequency of the crystal. The method consists in determining f_r and setting C at some value which makes f_1 and f_2 approximately symmetrical about f_r . The crystal inductance L_m is then found from the expression

$$L_m = \frac{f_1^2 f_2^2 L}{(f_r^2 - f_1^2)(f_2^2 - f_r^2)}$$

The true inductance, L , of the coil must be determined accurately and the effects of self-capacitance eliminated. This is done by making the coil resonate with two values of capacitance and measuring the frequency, the capacitor having been accurately calibrated.

The values of crystal inductance obtained by the two methods of measurement were found to agree to within 1%.

(4.2) Measurement of Equivalent Series Resistance

The apparatus used for measurement of series resistance is basically the same as that used for the measurement of equivalent inductance by the first method described. With the crystal inserted and the signal generator set to the resonant frequency, the output voltage of the generator is noted for a given reading on the detector. A short-circuit is applied across the crystal terminals and the output of the signal generator is adjusted to give the same reading on the detector. From the ratio of the two output voltages from the signal generator the crystal resistance can be calculated if the constants of the π -network are known. The same apparatus was also used for measuring the strength of unwanted resonances, although an automatic frequency-scanning method was also used for an approximate determination of the frequencies of unwanted modes.

(5) THE MEASURED PARAMETERS OF E.D.T. CRYSTALS

(5.1) The 0°-Y-cut

(5.1.1) Measurement of Unwanted Resonances.

As has been previously mentioned, more unwanted resonances are found in e.d.t. filter plates than in similar quartz crystals. It is therefore necessary to predict the frequency and strength of these resonances so as to be able to dimension the plates to avoid resonances close to the fundamental frequency of longitudinal vibration.

Design data were required for filters covering the frequency range 45–110 kc/s, and measurements were taken on over 250 specially made crystals of various dimensions, two batches of plates being made at frequencies of 60 and 100 kc/s. It was found that over this frequency range the frequency constants of unwanted modes were sensibly the same for crystals of similar width/length and length/thickness ratios.

It is to be expected that when a crystal with multiple electrodes is connected in a filter a component of the exciting field will be parallel to the major surfaces. In Y-cut e.d.t. crystals it has been found that such a field excites resonances not excited by a

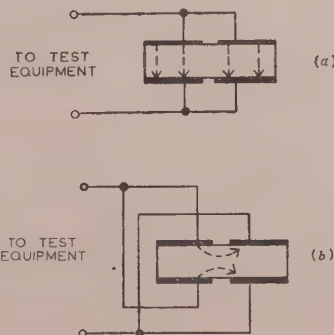


Fig. 6.—Connections for measuring unwanted resonances. (Sections of crystal perpendicular to length direction.)

(a) Direct.
(b) Transverse.

field normal to the major surfaces. Fig. 6 shows the two methods of electrode connection used in the measurement of spurious resonances. These are known as direct and transverse connections; they excite two series of resonances in Y-cut plates. With direct connections the unwanted modes are dependent on the width of the plate, but are independent of the plate thickness. With transverse connections the resonances are dependent on both width and thickness.

Three series of plates were made, having width/length ratios 0.35, 0.40 and 0.45, and from the results obtained intermediate values can be interpolated. Fig. 7 shows curves of the frequency

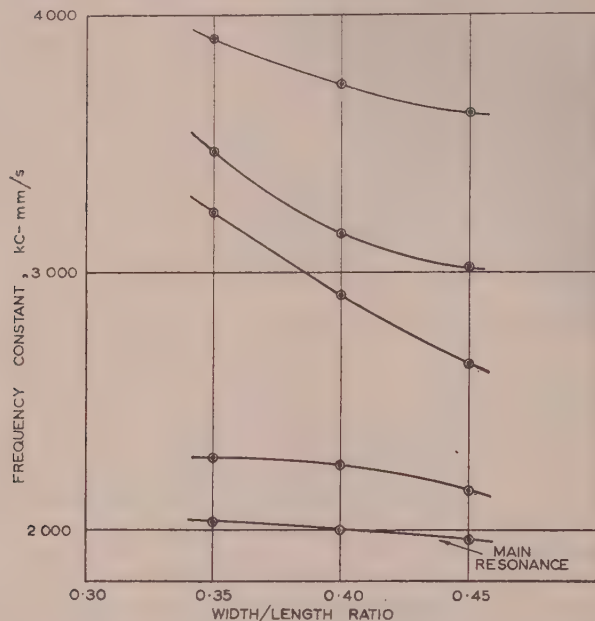


Fig. 7.—Resonance values with direct connections on Y-cut crystals of different width/length ratios.

constants of resonances obtained with direct connections, against width/length ratio. Although no unwanted resonances are found below the principal frequency, a well-defined series is situated above. The nearest of these unwanted resonances lies approximately 12% in frequency above the main response, but this is not sufficiently close to preclude the use of e.d.t. crystals in filter networks.

With transverse connections the resonances follow a different pattern, and may coincide in frequency with the main resonance. A typical resonance spectrum for crystals of width/length ratio 0.40 is shown in Fig. 8, the frequency constants being plotted against the length/thickness ratio. The direct-connection resonances are also shown on this Figure as straight lines parallel to the length/thickness-ratio axis. The curves for plates of width/length ratios 0.35 and 0.45 were found to be of similar form. The resonances obtained with transverse connections were also related to width/length ratio. A typical series of curves showing the relation between the frequency constant of one such resonance and the width/length ratio is shown in Fig. 9.

Because of the close proximity of the resonances obtained with transverse connections it is not always possible to obtain a usable crystal of given inductance by simply varying the dimensions of the plate. It was therefore necessary to employ partial plating to obtain the desired characteristics. This is discussed in detail in Section 6.

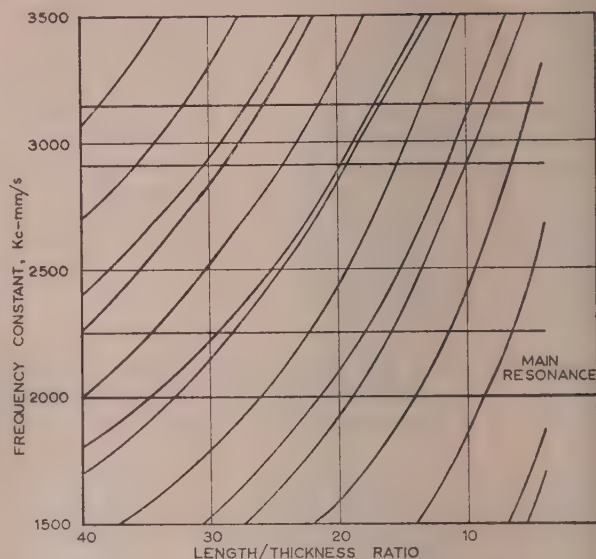


Fig. 8.—Resonance values with transverse connections on Y-cut crystals of width/length ratio 0.40 and different length/thickness ratios.

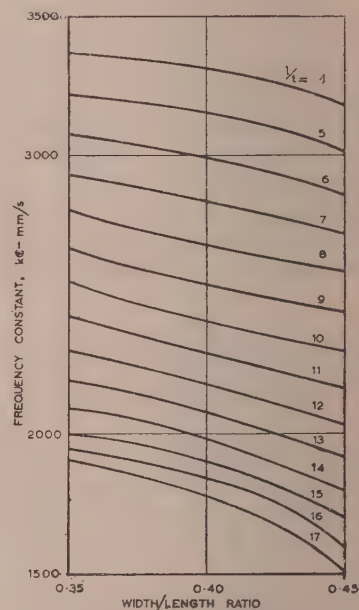


Fig. 9.—Variation in frequency constant of transverse resonance values for Y-cut crystals of different width/length and length/thickness ratios.

(5.1.2) Measurements of Parameters at the Principal Resonant Frequency.

Measurements were made of the frequency, equivalent inductance and equivalent series resistance at the main resonant frequency. The frequency constant of the main resonance is dependent on the width/length ratio but independent of length/thickness ratio, as may be seen from Figs. 7 and 8.

The inductance constant decreases slightly with increasing width/length ratio or increasing length/thickness ratio. A graph of inductance constant against width/length ratio is given in Fig. 10.

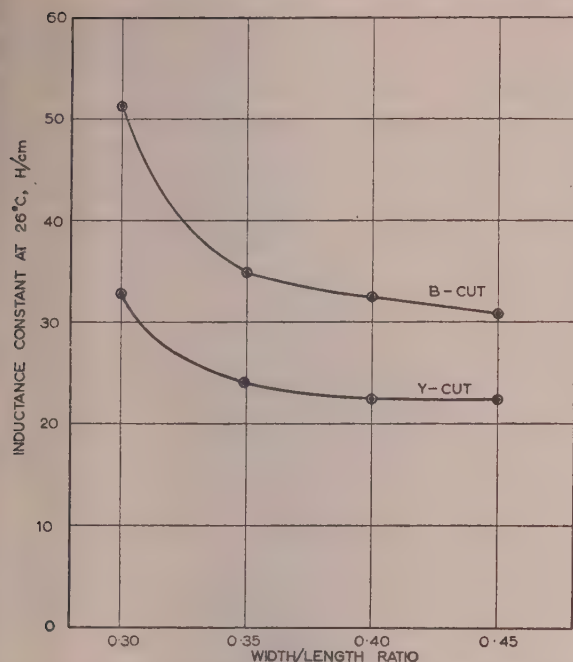


Fig. 10.—Variation in inductance constant with width/length ratio.

The equivalent series resistance R_m shows a considerable range due to fortuitous differences in mounting. It was thus not possible to determine practically a relation between series resistance and crystal dimensions.

5.1.3) Behaviour with varying Temperature.

The parameters at the principal resonant frequency were measured over the temperature range 15–45°C.

The frequency/temperature characteristic is parabolic, the inversion temperature being dependent on orientation and also to a smaller extent on dimensions. It is well known that the equation of such a curve may be expressed in the form

$$\frac{f_0 - f}{f_0} = A(T - T_0)^2$$

where f = Frequency at temperature T .

f_0 = Maximum frequency obtained at the inversion temperature T_0 (at vertex of parabola).

A = A constant, greater for greater variations of frequency with temperature.

It was found that A , known as the curvature constant, was practically independent of crystal dimensions.

The equivalent series inductance L_m is found to be related to temperature by the linear equation

$$L_m = L_{m0}(1 + \alpha T)$$

where L_{m0} = Inductance at 0°C.

α = Temperature coefficient of inductance, independent of crystal dimensions.

Measurements of equivalent series resistance were also made at various temperatures, and from these the Q-factor was calculated. For vacuum-mounted crystals the damping on the crystal is largely governed by the adhesive used in mounting. In general, a fall in the Q-factor occurs at high temperatures.

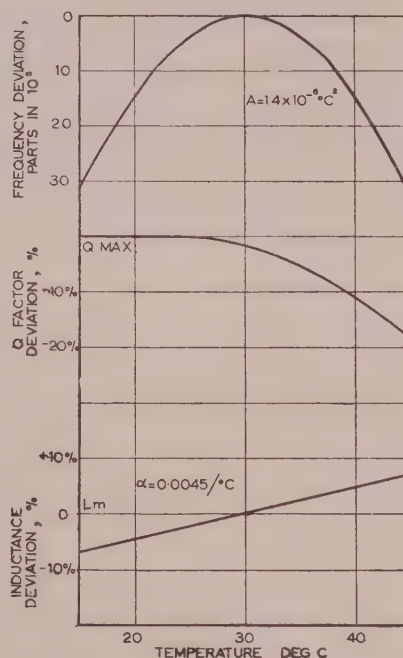


Fig. 11.—Effect of temperature on the parameters of Y-cut crystals.

Typical curves of inductance, frequency and Q-factor against temperature for a crystal mounted in vacuo are given in Fig. 11.

Mean values of the various constants are given in Table 1, which summarizes the parameters for various cuts. The values for an equivalent type of quartz plate are also given.

Table 1
CONSTANTS FOR VARIOUS CRYSTAL CUTS

Crystal	Width Length ratio	Frequency constant N	Curvature constant A per °C ²	Inductance constant K	Temperature coefficient of inductance α per °C	Inversion temperature T_0	Frequency difference of nearest unwanted resonance
E.D.T., 0° Y-cut	0.40	kc-mm/s	$\times 10^{-6}$	H/cm		°C	%
E.D.T., 0° Y-cut	0.40	1995	1.42	22.7	0.0045	29.6	11.3
E.D.T., B-cut	0.40	2053	1.26	32.5	0.0047	32.5	12.9
E.D.T., A-cut	0.25	1728	0.82	54.3	0.0039	21	7.9
Quartz, 18.5° X-cut	0.50	2545	—	133	—	—	34.2

(5.2) The B-cut

The frequency spectra of B-cut crystals are similar in form to those obtained with Y-cut plates. The two sets of curves differ, however, in that for B-cut crystals those resonances found with transverse connections are also found with direct connections. In general these responses are very small compared with the main response in both Y-cut and B-cut plates. Measurements were again made of inductance constant, frequency constant and equivalent series resistance. The temperature characteristics of inductance, frequency and equivalent series resistance have the same form as those of Y-cut crystals. Average values are given in Table 1.

(5.3) The A-cut

In A-cut crystals an unwanted resonance coincides with the fundamental response if the width/length ratio is approximately 0.35. The frequency constant of the A-cut is also smaller than that of the B- and Y-cuts. Experimental plates of width/length ratios 0.25, and 0.40 were constructed and only plates of the lower ratio showed sufficient freedom from unwanted resonances. Such crystals, however, are too narrow to permit the use of divided plating along their length. The resonance spectra and temperature characteristics of the A-cut were found to be similar in form to those of the B-cut. The measured constants are given in Table 1.

(5.4) High-Frequency Shear Mode

The orientation of the thickness shear mode is shown in Fig. 2(f). The frequency spectrum of these plates is extremely complex, and in order to obtain temperature-coefficient curves it was found necessary to use very large plates excited by small electrodes. The mean curvature constant of the frequency/temperature parabola for three 1 Mc/s plates was 0.97 compared with 0.04 for BT-cut quartz crystals. Because of the extreme difficulty in obtaining freedom from unwanted resonances and the fragility of the plates it was considered that the high-frequency shear mode was unsuitable for practical use.

(6) PARTIAL PLATING

In order to obtain crystals of a specified inductance free from unwanted resonances it is often necessary to use electrodes covering only part of the crystal surface. The electrodes are applied over the full width of the crystal but not over the whole length. It has been shown by Cady¹⁷ and others that the inductance L_p of a thin partially plated crystal operating in the longitudinal mode is found from the equation

$$\frac{L_F}{L_p} = \sin^2 \left(\frac{\pi l'}{2 l} \right)$$

where L_F = Inductance of a fully plated crystal of similar dimensions.
 l' = Length of the plated area.
 l = Length of the crystal.

This relation does not hold sufficiently accurately for Y- and B-cut crystals. Experimental results show inductances smaller than given by the equation. The extent of the deviation depends on both the length/thickness ratio and the orientation of the particular crystal. Fig. 12 shows some typical experimental results for both e.d.t. and quartz crystals compared with the theoretical curve.

(7) EFFECTS OF ACOUSTIC RADIATION ON ACTIVITY

When a low-frequency crystal is mounted in air a considerable radiation of energy occurs from the crystal. In a longitudinal mode X-cut quartz bar this radiation is mainly from the ends,

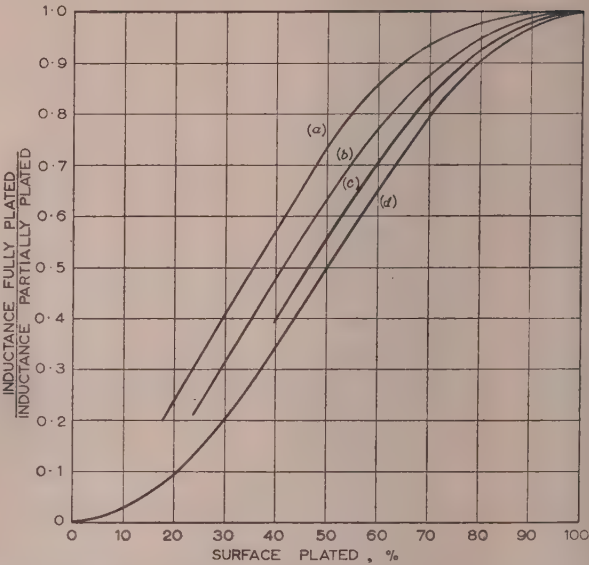


Fig. 12.—Effect of partial plating on inductance of Y-cut and B-cut crystals.

- (a) B-cut e.d.t.; length/thickness = 10 and 15.
- (b) Y-cut e.d.t.; length/thickness = 16 and 33.
- (c) 5° X-cut quartz; length/thickness = 33.
- (d) Theoretical curve of $L_F/L_p = \sin^2(\pi l'/2l)$.

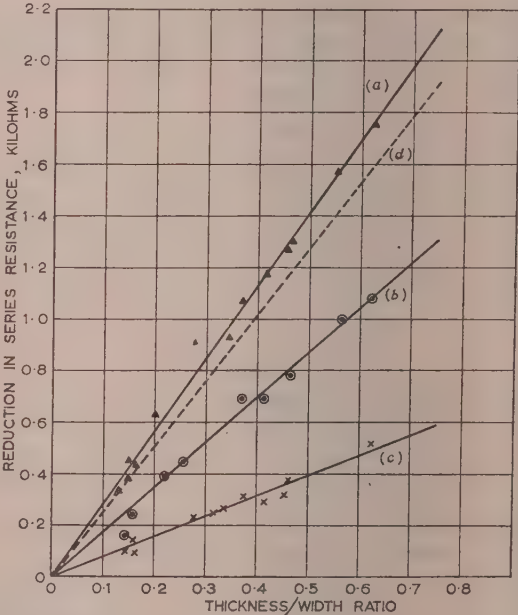


Fig. 13.—Effect on series resistance of Y-cut crystals of reductions in acoustic radiation losses from edges.

- (a) Losses reduced by evacuation.
- (b) Losses reduced by reflecting surfaces at $\lambda/4$ from short edges.
- (c) Losses reduced by reflecting surfaces at $\lambda/4$ from long edges.
- (d) Losses reduced by combined use of reflecting surfaces at $\lambda/4$ from short and long edges (broken line).

but for Y-cut e.d.t. crystals considerable radiation also occurs from the edges of the plate. If the radiation is reflected back to the crystal by some part of the crystal holder it can produce a considerable change in the activity of the plate, giving either a decrease or an increase, according to the phase of the reflected wave.

It is found that by placing reflectors at an odd number of quarter wavelengths from the crystal a considerable improvement in series resistance obtained with various combinations of reflectors, spaced at a quarter-wavelength from the edges of the crystals, compared with that produced by evacuation. For Y-cut crystals of width/length ratio 0.35 to 0.45 the reduction is proportional to the thickness/width ratio of the plate. It can be seen that, by placing reflectors round all the edges of the plates, activities very nearly as great as those obtained in vacuo can be attained.

(8) REFERENCES

- (1) WALKER, A. C., and KOHMAN, G. T.: 'Growing Crystals of Ethylene Diamine Tartrate', *Transactions of the American I.E.E.*, 1948, **67**, p. 565.
 - (2) WALKER, A. C.: 'Growing Piezoelectric Crystals', *Journal of the Franklin Institute*, 1950, **250**, p. 481.
 - (3) MASON, W. P.: 'New Low-Coefficient Synthetic Piezoelectric Crystals for use in Filters and Oscillators', *Proceedings of the Institute of Radio Engineers*, 1947, **35**, 10, p. 1005.
 - (4) GRIFFIN, J. P., and PENNELL, E. S.: 'Design and Performance of Ethylene Diamine Tartrate Crystal Units', *Transactions of the American I.E.E.*, 1948, **67**, p. 557.
 - (5) WILLIS, E. S.: 'Crystal Filters using Ethylene Diamine Tartrate in place of Quartz', *Transactions of the American I.E.E.*, 1948, **67**, p. 552.
 - (6) ZIEGLER, A. W.: 'Mechanical Development of EDT Crystal Units', *Bell Laboratories Record*, 1949, **27**, 7, p. 245.
 - (7) MASON, W. P.: 'Piezoelectric Crystals and their application to Ultrasonics' (Van Nostrand, 1950), p. 165.
 - (8) DAUNCEY, L. A.: 'Growing Ethylenediamine Tartrate Crystals', *The Times Review of Industry*, 1952, **6**, 27, p. 24.
 - (9) DAUNCEY, L. A.: 'The Solubility of Ethylenediamine d-tartrate in Water', *Nature*, 1951, **168**, p. 34.
 - (10) DAUNCEY, L. A., and STILL, J. E.: 'Apparatus for the Measurement of Saturation Temperature', *Journal of Applied Chemistry*, 1952, **2**, p. 399.
 - (11) British Patents Nos. 631345, 631348, 650796 and 655073.
 - (12) British Patents Nos. 656818 and 656819.
 - (13) British Patent No. 650795.
 - (14) British Patents Nos. 655994, 674895, 677898 and 679023.
 - (15) 'Standards on Piezoelectric Crystals', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, 12, p. 1378.
 - (16) HEISING, R. A.: 'Quartz Crystals for Electrical Circuits' (Van Nostrand, 1946), p. 279.
 - (17) CADY, W. G.: 'Piezoelectricity' (McGraw-Hill, 1946), p. 303.
 - (18) British Patent Application No. 21450/55.
-

LAGUERRE FUNCTIONS: TABLES AND PROPERTIES

By J. W. HEAD, M.A., and W. PROCTOR WILSON, C.B.E., B.Sc.(Eng.), Member.

(The paper was first received 14th January, and in revised form 5th March, 1956. It was published as an INSTITUTION MONOGRAPH in June, 1956.)

SUMMARY

The first 21 Laguerre functions have been tabulated to four decimal places (or two significant figures when this is more accurate) for values of the argument $0(0.1)1(0.2)3(0.5)6(1)14(2)40(5)100$. The practical applications of these functions are briefly considered. The zeros of these Laguerre functions are also tabulated to 5 places, the last of which is not reliable except in the case of the smallest zero for orders above 6. Differentiation and integration of Laguerre functions, orthonormal properties and addition formulae are briefly discussed. Finite or infinite series of Laguerre functions whose sums are well-known functions such as products of polynomials and circular or exponential functions, step and delta functions, Bessel and Kelvin functions, and gamma and error functions are stated. The general methods used for deriving some of these results are indicated in Section 13.

LIST OF SYMBOLS

The symbols listed below occur frequently and have been assumed to be well known. All symbols not in this list are defined where they appear, and occur only occasionally.

$H(x)$ = Heaviside unit function.

$I_0(x)$, $I_1(x)$ = Bessel functions of imaginary argument.

$J_0(x)$ = Bessel function (of the first kind).

$L_n(x)$ = Laguerre polynomial of degree n .

$\binom{n}{r}$ = Binomial coefficient $\frac{n!}{(n-r)!r!}$.

p = Heaviside operator.

[If two functions $F(x)$ and $F[p]$ are operationally equivalent, i.e. related by the Carson integral, see eqn. (60), the argument is placed in round brackets in the x -domain and in square brackets in the p -domain with the same symbol F .]

$P_n(x)$ = Legendre polynomial of order n .

$T_n(x)$, $U_n(x)$ = Chebyshev* function of order n .

$[T_n(x) = \cos(n \cos^{-1} x); U_n(x) = \sin(n \cos^{-1} x).]$

α_{nr} = r th zero of $L_n(x)$ in ascending order.

$\Gamma(x)$ = Gamma function.

$\gamma(n, x)$ = Incomplete gamma function.

$\delta(x)$ = Dirac delta function.

$\lambda_n(x)$ = Laguerre function of order n , $\varepsilon^{-x/2}L_n(x)$.

(1) INTRODUCTION TO TABLES

We define the Laguerre function $\lambda_n(x)$ † by the equation

$$\lambda_n(x) = \varepsilon^{-x/2}L_n(x) \quad (1)$$

where $L_n(x)$ is the Laguerre polynomial of order n given by

$$L_n(x) = 1 - \binom{n}{1}x + \binom{n}{2}\frac{x^2}{2!} - \dots + (-1)^n \frac{x^n}{n!} \quad (2)$$

* An alternative spelling is Tscheycheff.

† Jahnke and Emde (Reference 1, pp. 32, 33) use the notation $l_n(x)$ in their tables of this function.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Head and Mr. Wilson are with the British Broadcasting Corporation.

$\binom{n}{r}$ being the binomial coefficient

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad (3)$$

The Tables give $\lambda_n(x)$ for $n = 0(1)20$ and for $x = 0(0.1)1(0.2)3(0.5)6(1)14(2)40(5)100$. Values are computed to four places of decimals except when $|\lambda_n(x)| < 10^{-3}$ and there is no zero of $\lambda_n(x)$ in the immediate vicinity; in this case the values are given to two significant figures.

The recurrence relation between successive Laguerre functions is

$$(n+1)\lambda_{n+1}(x) = (2n+1-x)\lambda_n(x) - n\lambda_{n-1}(x) \quad (4)$$

while $\varepsilon^{-kx}\lambda_n(x)$ satisfies the differential equation

$$x \frac{d^2z}{dx^2} + (1+2kx) \frac{dz}{dx} + [(k^2 - \frac{1}{4})x + (n+k+\frac{1}{2})]z = 0 \quad (5)$$

for all values of n , k and x .

If $k = -\frac{1}{2}$, (5) becomes the differential equation satisfied by Laguerre polynomials. This last equation can also be derived by regarding Laguerre polynomials as a special kind of confluent hypergeometric function.

For numerical work with the Tables which follow, the obvious identity

$$L_n(x) \equiv \lambda_n(x)/\lambda_0(x)$$

is useful.

The Laguerre functions have many applications to transient phenomena in electrical networks and dynamical systems. The Fourier transforms (and the allied operational images) of these functions correspond to transfer impedance or admittance functions of physically realizable networks. Lee,² Wiener³ and others have made use of these transforms for network synthesis.

There are certain repeated networks having a response to unit-function excitation naturally expressed in terms of Laguerre functions. Such networks have been studied by Wagner,⁴ van der Pol and Bremmer,⁵ and by the authors (see Section 13.6). In a recent Monograph, Lampard⁶ has described two examples of 'Laguerre function filters'.

(2) ZEROS OF LAGUERRE FUNCTIONS $\lambda_n(x)$ FOR $n = 1, 2, \dots, 20$

Table 5 gives the values of the 210 zeros of the Laguerre functions $\lambda_n(x)$ for $n = 1, 2, \dots, 20$. They are given to five decimal places, but the fifth place should not be regarded as reliable, except in the case of the smallest zero for $n > 7$.

The method of calculation is discussed briefly in Section 13.1.

Since

$$L_n(k/n) = 1 - \frac{k}{(1!)^2} + \frac{k^2}{(2!)^2} \left(1 - \frac{1}{n}\right) - \frac{k^3}{(3!)^2} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) + \dots \quad (6)$$

we expect at any rate the smaller zeros of $L_n(x)$ to differ little from those of the Bessel function

$$J_0[2\sqrt{(nx)}] = J_0(2\sqrt{k}) = 1 - \frac{k}{(1!)^2} + \frac{k^2}{(2!)^2} - \frac{k^3}{(3!)^2} + \dots \quad (7)$$

Table 1

LAGUERRE FUNCTIONS $\lambda_n(x)$ $x = 0$ to 10; $n = 0$ to 10

x	$\lambda_0(x)$	$\lambda_1(x)$	$\lambda_2(x)$	$\lambda_3(x)$	$\lambda_4(x)$	$\lambda_5(x)$	$\lambda_6(x)$	$\lambda_7(x)$	$\lambda_8(x)$	$\lambda_9(x)$	$\lambda_{10}(x)$
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	0.9512	0.8561	0.7657	0.6799	0.5986	0.5215	0.4489	0.3801	0.3154	0.2544	0.1970
0.2	0.9048	0.7238	0.5610	0.4150	0.2847	0.1692	0.0672	-0.0221	-0.0997	-0.1665	-0.2232
0.3	0.8607	0.6025	0.3830	0.1984	0.0450	-0.0803	-0.1808	-0.2592	-0.3180	-0.3598	-0.3865
0.4	0.8187	0.4912	0.2292	0.0240	-0.1323	-0.2468	-0.3257	-0.3747	-0.3990	-0.4018	-0.3883
0.5	0.7788	0.3894	0.0973	-0.1135	-0.2575	-0.3470	-0.3927	-0.4037	-0.3882	-0.3528	-0.3033
0.6	0.7408	0.2963	-0.0148	-0.2193	-0.3397	-0.3953	-0.4021	-0.3735	-0.3204	-0.2519	-0.1751
0.7	0.7047	0.2114	-0.1092	-0.2975	-0.3866	-0.4038	-0.3710	-0.3058	-0.2220	-0.1302	-0.0385
0.8	0.6703	0.1341	-0.1877	-0.3521	-0.4051	-0.3825	-0.3128	-0.2173	-0.1120	-0.0084	0.0855
0.9	0.6376	0.0638	-0.2519	-0.3867	-0.4008	-0.3400	-0.2383	-0.1204	-0.0038	0.1002	0.1848
1.0	0.6065	0.0000	-0.3033	-0.4043	-0.3791	-0.2830	-0.1558	-0.0245	0.0934	0.1879	0.2541
1.2	0.5488	-0.1098	-0.3732	-0.3995	-0.2994	-0.1475	-0.0086	0.1410	0.2356	0.2883	0.3012
1.4	0.4966	-0.1986	-0.4072	-0.3562	-0.1933	0.0088	0.1469	0.2511	0.2983	0.2938	0.2487
1.6	0.4493	-0.2696	-0.4134	-0.2888	-0.0798	0.1129	0.2434	0.2996	0.2889	0.2280	0.1367
1.8	0.4066	-0.3253	-0.3984	-0.2082	0.0282	0.2072	0.2941	0.2930	0.2262	0.1215	0.0054
2.0	0.3679	-0.3679	-0.3679	-0.1226	0.1226	0.2698	0.3025	0.2441	0.1320	0.0030	-0.1137
2.2	0.3329	-0.3994	-0.3262	-0.0382	0.1989	0.3010	0.2757	0.1674	0.0266	-0.1051	-0.2004
2.4	0.3012	-0.4217	-0.2771	0.0410	0.2549	0.3037	0.2229	0.0772	-0.0734	-0.1878	-0.2456
2.6	0.2725	-0.4361	-0.2235	0.1119	0.2907	0.2826	0.1534	-0.0144	-0.1565	-0.2376	-0.2488
2.8	0.2466	-0.4439	-0.1677	0.1729	0.3074	0.2428	0.0756	-0.0979	-0.2154	-0.2529	-0.2158
3.0	0.2231	-0.4463	-0.1116	0.2231	0.3068	0.1897	-0.0028	-0.1665	-0.2474	-0.2368	-0.1562
3.5	0.1738	-0.4344	0.0217	0.3005	0.2466	0.0309	-0.1669	-0.2530	-0.2176	-0.1016	0.0384
4.0	0.1353	-0.4060	0.1353	0.3158	0.1353	-0.1173	-0.2496	-0.2204	-0.0846	0.0737	0.1867
4.5	0.1054	-0.3689	0.2240	0.2833	0.0091	-0.2185	-0.2442	-0.1093	0.0702	0.1947	0.2191
5.0	0.0821	-0.3283	0.2873	0.2189	-0.1060	-0.2599	-0.1716	0.0267	0.1835	0.2209	0.1442
5.5	0.0639	-0.2877	0.3276	0.1372	-0.1943	-0.2457	-0.0634	0.1427	0.2249	0.1605	+0.0143
6	0.0498	-0.2489	0.3485	0.0498	-0.2489	-0.1892	0.0498	0.2119	0.1949	0.0498	-0.1107
7	0.0302	-0.1812	0.3473	-0.1107	-0.2604	-0.0156	0.2066	0.1905	0.0097	-0.1586	-0.1990
8	0.0183	-0.1282	0.3114	-0.2259	-0.1770	0.1453	0.2202	0.0327	-0.1640	-0.1931	-0.0648
9	0.0111	-0.0889	0.2611	-0.2888	-0.0514	0.2311	0.1198	-0.1296	-0.2020	-0.0644	0.1174
10	0.0067	-0.0606	0.2089	-0.3077	0.0741	0.2313	-0.0232	-0.2082	-0.1098	0.0997	0.1886

The first eight zeros of (7) occur when

$$k = nx = 1.4458, 7.6178, 18.7217, 34.7600, 55.7332, \\ 81.6408, 112.4834 \text{ and } 148.2607. \quad (8)$$

and by comparing (8) for various n with the tabulated zeros of $L_n(x)$, we find that the correspondence between Laguerre and Bessel zeros is remarkably close unless we are considering Laguerre zeros of relatively large modulus for the order (value of n) under discussion. The zeros can be used to determine an expansion of a function by means of a Laguerre-function series; this is discussed more fully below. A profile diagram showing the general behaviour of Laguerre functions of order not exceeding 10 is given in Fig. 1.

(3) NOTATIONS AND DEFINITIONS

The following notations for and definitions of the Laguerre polynomial are or have been in use in the literature:

(a) Laguerre's original notation.⁷

$$f_n(x) = \sum_{s=0}^n \frac{n^2(n-1)^2 \dots (s+1)^2}{(n-s)!} x^s \quad (9)$$

Thus Laguerre's polynomial $f_n(x)$ is our $L_n(-x)$ multiplied by $n!$, according to (2).

(b) Notation of van der Pol, Madelung, *et al.*

$$L_n(x) = \sum_{s=0}^n (-1)^s \frac{n^2(n-1)^2 \dots (s+1)^2}{(n-s)!} x^s = \varepsilon^x \frac{d^n}{dx^n} (\varepsilon^{-x} x^n) \quad (10)$$

so that the polynomial $L_n(x)$ given by (10) is $n!$ times our $L_n(x)$ given by (2).

(c) Notation of Wiener.³

$$\varepsilon^x L_n(x) = \sqrt{2} \sum_{s=0}^n (-1)^{n+s} \frac{n!}{(n-s)!(s!)^2} (2x)^s \quad (11)$$

so that $L_n(x)$ given by (11) is, according to our definition (1),

$$(-1)^n \sqrt{2} \lambda_n(2x)$$

and is thus a Laguerre function, not a polynomial.

(d) Notation of Wagner, Tricomi, Doetsch, MacLachlan, *et al.*, which is the notation here adopted.

$$L_n(x) = \sum_{s=0}^n (-1)^s \frac{n!}{(n-s)!(s!)^2} x^s = \frac{\varepsilon^x}{n!} \frac{d^n}{dx^n} (\varepsilon^{-x} x^n) \\ = \frac{1}{n!} \left(\frac{d}{dx} - 1 \right)^n x^n = {}_1F_1(-n; 1; x)^* \quad (12)$$

the last member of (12) being a confluent hypergeometric function. This is the same as (2), and the definition (1) of the Laguerre function is also used by these authors.

* This function is attributed to Kummer by some writers, with a different notation. Abel's polynomial, designated $\Phi(x)$ by French writers, also coincides with the definition here adopted for $L_n(x)$.

Table 2
LAGUERRE FUNCTIONS $\lambda_n(x)$
 $x = 0$ to 10 ; $n = 11$ to 20

x	$\lambda_{11}(x)$	$\lambda_{12}(x)$	$\lambda_{13}(x)$	$\lambda_{14}(x)$	$\lambda_{15}(x)$	$\lambda_{16}(x)$	$\lambda_{17}(x)$	$\lambda_{18}(x)$	$\lambda_{19}(x)$	$\lambda_{20}(x)$
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.1	0.1416	0.0907	0.0430	-0.0016	-0.0432	-0.0819	-0.1179	-0.1512	-0.1820	-0.2103
0.2	-0.2707	-0.3098	-0.3411	-0.3653	-0.3830	-0.3948	-0.4012	-0.4029	-0.4002	-0.3937
0.3	-0.4003	-0.4029	-0.3960	-0.3811	-0.3596	-0.3327	-0.3015	-0.2671	-0.2302	-0.1917
0.4	-0.3643	-0.3285	-0.2854	-0.2372	-0.1859	-0.1332	-0.0804	-0.0288	0.0207	0.0673
0.5	-0.2445	-0.1804	-0.1143	-0.0489	0.0138	0.0722	0.1250	0.1714	0.2108	0.2430
0.6	-0.0957	-0.0182	0.0542	0.1191	0.1750	0.2207	0.2560	0.2808	0.2954	0.3005
0.7	0.0473	0.1232	0.1867	0.2362	0.2715	0.2926	0.3005	0.2963	0.2813	0.2573
0.8	0.1646	0.2262	0.2691	0.2936	0.3007	0.2924	0.2709	0.2384	0.1977	0.1511
0.9	0.2467	0.2849	0.3004	0.2955	0.2731	0.2369	0.1902	0.1366	0.0793	0.0214
1.0	0.2912	0.3010	0.2868	0.2532	0.2049	0.1469	0.0836	0.0192	-0.0428	-0.0996
1.2	0.2800	0.2326	0.1673	0.0924	0.0151	-0.0585	-0.1237	-0.1770	-0.2163	-0.2407
1.4	0.1760	0.0888	-0.0012	-0.0847	-0.1547	-0.2068	-0.2388	-0.2504	-0.2430	-0.2189
1.6	0.0338	-0.0650	-0.1482	-0.2085	-0.2426	-0.2503	-0.2339	-0.1977	-0.1468	-0.0866
1.8	-0.1010	-0.1834	-0.2341	-0.2510	-0.2367	-0.1967	-0.1382	-0.0691	0.0029	0.0710
2.0	-0.1991	-0.2442	-0.2483	-0.2166	-0.1581	-0.0835	-0.0035	0.0724	0.1368	0.1842
2.2	-0.2471	-0.2445	-0.2008	-0.1286	-0.0424	0.0443	0.1201	0.1770	0.2105	0.2191
2.4	-0.2446	-0.1947	-0.1128	-0.0173	0.0745	0.1495	0.1989	0.2191	0.2105	0.1771
2.6	-0.2002	-0.1123	-0.0087	0.0892	0.1650	0.2093	0.2190	0.1965	0.1483	0.0832
2.8	-0.1272	-0.0162	0.0897	0.1701	0.2134	0.2166	0.1840	0.1246	0.0499	-0.0280
3.0	-0.0403	0.0759	0.1658	0.2136	0.2156	0.1770	0.1095	0.0274	-0.0546	-0.1244
3.5	0.1535	0.2142	0.2125	0.1579	0.0700	-0.0276	-0.1139	-0.1732	-0.1975	-0.1860
4.0	0.2215	0.1796	0.0857	-0.0260	-0.1233	-0.1837	-0.1974	-0.1664	-0.1020	-0.0204
4.5	0.1517	0.0330	-0.0880	-0.1721	-0.1989	-0.1681	-0.0946	-0.0016	0.0870	0.1515
5.0	0.0088	-0.1189	-0.1911	-0.1899	-0.1254	-0.0258	0.0755	0.1502	0.1815	0.1658
5.5	-0.1258	-0.1966	-0.1787	-0.0919	0.0228	0.1225	0.1767	0.1739	0.1209	0.0373
6	-0.1962	-0.1765	-0.0768	0.0486	0.1463	0.1829	0.1529	0.0736	-0.0248	-0.1109
7	-0.1091	0.0369	0.1519	0.1826	0.1261	0.0180	-0.0912	-0.1589	-0.1644	-0.1122
8	0.0990	0.1831	0.1481	0.0310	-0.0949	-0.1654	-0.1540	-0.0747	0.0318	0.1203
9	0.1867	0.1101	-0.0368	-0.1495	-0.1651	-0.0868	0.0328	0.1294	0.1596	0.1164
10	0.0979	-0.0667	-0.1674	-0.1413	-0.0228	0.1026	0.1603	0.1257	0.0267	-0.0806

(4) ORTHONORMAL PROPERTIES

Henceforward we shall only use the notation of (1), (2) and (12). The Laguerre functions have orthonormal properties

$$\int_0^\infty \lambda_m(x) \lambda_n(x) dx = \begin{cases} 0; & (m \neq n) \\ 1; & (m = n) \end{cases} \quad (13)$$

Now suppose that we have a function $f(x)$ of x which is negligible for $x > x_0$; let us attempt to express it as a Laguerre series

$$f(x) = \varepsilon^{\mu x} \sum_{s=0}^N \alpha_s \lambda_s(kx) \quad (14)$$

where μ, k are real but otherwise arbitrary.

We then find, by using (13),

$$\alpha_s = k \int_0^\infty \varepsilon^{-\mu x} f(x) \lambda_s(kx) dx \quad (15)$$

Now (14) will not be a satisfactory representation unless $\mu < \frac{1}{2}k$, for if $\mu \geq \frac{1}{2}k$, the right-hand side of (14) is large for large x , contrary to the hypothesis $f(x)$ is negligible for $x > x_0$. Again, if $\mu < -\frac{1}{2}k$, the expression $\varepsilon^{-\mu x} \lambda_s(kx)$ would contain an exponential with positive exponent which could prevent us from justifiably replacing the upper limit of integration in (15) by x_0 . Hence in the absence of further information about the

behaviour of $f(x)$ for $x > x_0$ it is inconvenient to take μ, k except in such a way that $-\frac{1}{2}k < \mu < \frac{1}{2}k$. If the approximation (14) is to be useful it is clearly desirable that α_s should be negligible when s exceeds a reasonably small value s_0 if μ, k can be chosen suitably. If this can be done, the effective value of N in (14) is s_0 , otherwise the right-hand side of (14) is an infinite series. The proper choice of μ, k is discussed below for the case when $f(x)$ is given graphically and in Section 13.3 when $f(x)$ is of the form

$$f(x) = \sum A_s \varepsilon^{-\alpha_s x} \quad (16)$$

where the A 's and α 's may be real or may form complex conjugate pairs.

If we know that $f(x)$ can be expressed in the form

$$f(x) = \sum_{s=0}^n \alpha_s(kx)^s \varepsilon^{-1/2kx} / s! \quad (17)^*$$

then (67) shows that

$$\begin{aligned} f(x) = & \left(\sum_{s=0}^n \alpha_s \right) \lambda_0(kx) - \left(\sum_{s=1}^n s \alpha_s \right) \lambda_1(kx) + \left[\sum_{s=2}^n \frac{s(s-1)}{2!} \alpha_s \right] \lambda_2(kx) \\ & - \dots + \left[\sum_{s=r}^n (-1)^{s(r-1)} \frac{s(s-1) \dots (s-r+1)}{r!} \alpha_s \right] \lambda_r(kx) + \dots \\ & + (-1)^n \alpha_n \lambda_n(kx) \quad (18) \end{aligned}$$

* Throughout the paper, an expression of the form $\varepsilon^{-1/2kx}$ is to be interpreted as $\varepsilon^{-kx/2}$ not $\varepsilon^{-1/(2kx)}$.

Table 3
LAGUERRE FUNCTIONS $\lambda_n(x)$
 $x = 11$ to 100; $n = 0$ to 10

$\lambda_0(x)$	$\lambda_1(x)$	$\lambda_2(x)$	$\lambda_3(x)$	$\lambda_4(x)$	$\lambda_5(x)$	$\lambda_6(x)$	$\lambda_7(x)$	$\lambda_8(x)$	$\lambda_9(x)$	$\lambda_{10}(x)$
0.0041	-0.0409	0.1614	-0.2956	0.1745	0.1667	-0.1454	-0.1844	0.0351	0.1873	0.1183
0.0025	-0.0273	0.1215	-0.2652	0.2404	0.0679	-0.2117	-0.0884	0.1520	0.1631	-0.0227
0.0015	-0.0180	0.0894	-0.2265	0.2727	-0.0369	-0.2149	0.0317	0.1960	0.0590	-0.1410
9.1×10^{-4}	-0.0118	0.0647	-0.1863	0.2775	-0.1285	-0.1670	0.1340	0.1629	-0.0648	-0.1790
3.3×10^{-4}	-0.0050	0.0325	-0.1160	0.2365	-0.2383	0.0015	0.2036	-0.0268	-0.1840	-0.0311
1.2×10^{-4}	-0.0021	0.0157	-0.0665	0.1712	-0.2549	0.1547	0.1080	-0.1759	-0.0764	0.1506
4.5×10^{-5}	-8.6 $\times 10^{-4}$	0.0073	-0.0360	0.1115	-0.2165	0.2318	-0.0463	-0.1739	0.0991	0.1466
1.7×10^{-5}	-3.5 $\times 10^{-4}$	0.0033	-0.0186	0.0673	-0.1600	0.2373	-0.1679	-0.0607	0.1830	-0.0003
6.1×10^{-6}	-1.4 $\times 10^{-4}$	0.0015	-0.0093	0.0383	-0.1076	0.2012	-0.2239	0.0759	0.1400	-0.1383
2.3×10^{-6}	-5.7 $\times 10^{-5}$	6.5 $\times 10^{-4}$	-0.0045	0.0209	-0.0675	0.1512	-0.2230	0.1743	0.0239	-0.1736
8.3×10^{-7}	-2.2 $\times 10^{-5}$	2.8 $\times 10^{-4}$	-0.0021	0.0110	-0.0401	0.1044	-0.1893	0.2163	-0.0961	-0.1082
3.1×10^{-7}	-8.9 $\times 10^{-6}$	1.2 $\times 10^{-4}$	-0.0010	0.0056	-0.0228	0.0674	-0.1441	0.2113	-0.1771	0.0046
1.1×10^{-7}	-3.5 $\times 10^{-6}$	5.1 $\times 10^{-5}$	-4.5 $\times 10^{-4}$	0.0028	-0.0125	0.0415	-0.1018	0.1801	-0.2097	0.1105
4.1×10^{-8}	-1.4 $\times 10^{-6}$	2.1 $\times 10^{-5}$	-2.0 $\times 10^{-4}$	0.0013	-0.0066	0.0241	-0.0667	0.1372	-0.2000	0.1764
1.5×10^{-8}	-5.3 $\times 10^{-7}$	8.8 $\times 10^{-6}$	-9.1 $\times 10^{-5}$	6.5 $\times 10^{-4}$	-0.0034	0.0138	-0.0423	0.0989	-0.1713	0.2022
5.6×10^{-9}	-2.1 $\times 10^{-7}$	3.6 $\times 10^{-6}$	-4.0 $\times 10^{-5}$	3.1 $\times 10^{-4}$	-0.0017	0.0076	-0.0255	0.0668	-0.1332	0.1929
2.0×10^{-9}	-8.0 $\times 10^{-8}$	1.5 $\times 10^{-6}$	-1.7 $\times 10^{-5}$	1.4 $\times 10^{-4}$	-8.6 $\times 10^{-4}$	0.0041	-0.0149	0.0430	-0.0967	0.1643
1.7×10^{-10}	-7.4 $\times 10^{-9}$	1.6 $\times 10^{-7}$	-2.1 $\times 10^{-6}$	2.0 $\times 10^{-5}$	-1.4 $\times 10^{-4}$	7.75 $\times 10^{-4}$	-0.0034	0.0122	-0.0348	0.0795
1.4×10^{-11}	-6.8 $\times 10^{-10}$	1.6 $\times 10^{-8}$	-2.4 $\times 10^{-7}$	2.6 $\times 10^{-6}$	-2.1 $\times 10^{-5}$	1.3 $\times 10^{-4}$	-6.9 $\times 10^{-4}$	0.0029	-0.0100	0.0283
1.1×10^{-12}	-6.2 $\times 10^{-11}$	1.6 $\times 10^{-9}$	-2.7 $\times 10^{-8}$	3.2 $\times 10^{-7}$	-2.9 $\times 10^{-6}$	2.1 $\times 10^{-5}$	-1.2 $\times 10^{-4}$	6.0 $\times 10^{-4}$	-0.0024	0.0082
9.3×10^{-14}	-5.5 $\times 10^{-12}$	1.6 $\times 10^{-10}$	-2.9 $\times 10^{-9}$	3.8 $\times 10^{-8}$	-3.9 $\times 10^{-7}$	3.1 $\times 10^{-6}$	-2.1 $\times 10^{-5}$	1.1 $\times 10^{-4}$	-5.2 $\times 10^{-4}$	0.0020
7.7×10^{-15}	-4.9 $\times 10^{-13}$	1.5 $\times 10^{-11}$	-3.0 $\times 10^{-10}$	4.4 $\times 10^{-9}$	-4.9 $\times 10^{-8}$	4.4 $\times 10^{-7}$	-3.2 $\times 10^{-6}$	2.0 $\times 10^{-5}$	-1.0 $\times 10^{-4}$	4.5 $\times 10^{-4}$
6.3×10^{-16}	-4.3 $\times 10^{-14}$	1.5 $\times 10^{-12}$	-3.2 $\times 10^{-11}$	5.0 $\times 10^{-10}$	-6.0 $\times 10^{-9}$	5.9 $\times 10^{-8}$	-4.7 $\times 10^{-7}$	3.2 $\times 10^{-6}$	-1.8 $\times 10^{-5}$	9.1 $\times 10^{-5}$
5.2×10^{-17}	-3.8 $\times 10^{-15}$	1.4 $\times 10^{-13}$	-3.2 $\times 10^{-12}$	5.5 $\times 10^{-11}$	-7.2 $\times 10^{-10}$	7.6 $\times 10^{-9}$	-6.7 $\times 10^{-8}$	4.9 $\times 10^{-7}$	-3.1 $\times 10^{-6}$	1.7 $\times 10^{-5}$
4.2×10^{-18}	-3.4 $\times 10^{-16}$	1.3 $\times 10^{-14}$	-3.2 $\times 10^{-13}$	5.9 $\times 10^{-12}$	-8.3 $\times 10^{-11}$	9.5 $\times 10^{-10}$	-9.0 $\times 10^{-9}$	7.3 $\times 10^{-8}$	-5.0 $\times 10^{-7}$	3.0 $\times 10^{-6}$
3.4×10^{-19}	-2.9 $\times 10^{-17}$	1.2 $\times 10^{-15}$	-3.2 $\times 10^{-14}$	6.2 $\times 10^{-13}$	-9.4 $\times 10^{-12}$	1.2 $\times 10^{-10}$	-1.2 $\times 10^{-9}$	1.0 $\times 10^{-8}$	-7.7 $\times 10^{-8}$	5.0 $\times 10^{-7}$
2.9×10^{-20}	-2.5 $\times 10^{-18}$	1.1 $\times 10^{-16}$	-3.1 $\times 10^{-15}$	6.5 $\times 10^{-14}$	-1.1 $\times 10^{-12}$	1.4 $\times 10^{-11}$	-1.5 $\times 10^{-10}$	1.4 $\times 10^{-9}$	-1.1 $\times 10^{-8}$	7.8 $\times 10^{-8}$
2.3×10^{-21}	-2.2 $\times 10^{-19}$	1.0 $\times 10^{-17}$	-3.0 $\times 10^{-16}$	6.7 $\times 10^{-15}$	-1.1 $\times 10^{-13}$	1.6 $\times 10^{-12}$	-1.9 $\times 10^{-11}$	1.9 $\times 10^{-10}$	-1.6 $\times 10^{-9}$	1.2 $\times 10^{-8}$
1.9×10^{-22}	-1.9 $\times 10^{-20}$	9.2 $\times 10^{-19}$	-2.9 $\times 10^{-17}$	6.8 $\times 10^{-16}$	-1.2 $\times 10^{-14}$	1.8 $\times 10^{-13}$	-2.3 $\times 10^{-12}$	2.4 $\times 10^{-11}$	-2.2 $\times 10^{-10}$	1.7 $\times 10^{-9}$

Table 4
LAGUERRE FUNCTIONS $\lambda_n(x)$
 $x = 11$ to 100; $n = 11$ to 20

x	$\lambda_{11}(x)$	$\lambda_{12}(x)$	$\lambda_{13}(x)$	$\lambda_{14}(x)$	$\lambda_{15}(x)$	$\lambda_{16}(x)$	$\lambda_{17}(x)$	$\lambda_{18}(x)$	$\lambda_{19}(x)$	$\lambda_{20}(x)$
11	-0.0627	-0.1712	-0.1264	+0.0145	0.1353	0.1556	0.0740	-0.0483	-0.1362	-0.1448
12	-0.1668	-0.1321	0.0219	0.1461	0.1452	0.0354	-0.0929	-0.1522	-0.1122	-0.0069
13	-0.1561	-0.0009	0.1433	0.1441	0.0200	-0.1127	-0.1513	-0.0786	0.0441	0.1320
14	-0.0550	0.1228	0.1547	0.0296	-0.1148	-0.1497	-0.0593	0.0722	0.1436	0.1109
16	0.1531	0.1178	-0.0598	-0.1564	-0.0797	0.0719	0.1469	0.0872	-0.0428	-0.1320
18	0.1106	-0.0920	-0.1516	-0.0120	0.1327	0.1191	-0.0198	-0.1312	0.0666	0.0066
20	-0.0767	-0.1535	0.0118	0.1484	0.0781	-0.0855	-0.1388	-0.0350	0.1002	0.1285
22	-0.1663	-0.0136	0.1504	0.0663	-0.1094	-0.1231	0.0229	0.1334	0.0836	-0.0557
24	-0.0896	0.1342	0.0930	-0.1047	-0.1217	0.0449	0.1384	0.0421	-0.1023	-0.1167
26	0.0572	0.1449	-0.0640	-0.1391	0.0319	0.1404	0.0278	-0.1187	-0.0951	0.0510
28	0.1561	0.0341	-0.1520	-0.0208	0.1405	0.0458	-0.1187	-0.0895	0.0701	0.1235
30	0.1572	-0.0960	-0.1082	0.1123	0.0935	-0.0994	-0.1056	0.0647	0.1238	-0.0056
32	0.0799	-0.1608	0.0129	0.1447	-0.0410	-0.1331	0.0307	0.1308	0.0053	-0.1224
34	-0.0269	-0.1382	0.1206	0.0681	-0.1352	-0.0385	0.1295	0.0435	-0.1158	-0.0703
36	-0.1199	-0.0544	0.1576	-0.0499	-0.1238	0.0854	0.1014	-0.0863	-0.1006	0.0669
38	-0.1771	0.0445	0.1190	-0.1348	-0.0302	0.1396	-0.0127	-0.1297	0.0188	0.1242
40	-0.1959	0.1269	0.0344	-0.1498	0.0778	0.0967	-0.1130	-0.0599	0.1165	0.0511
45	-0.1418	0.1871	-0.1570	0.0281	0.1166	-0.1283	-0.0191	0.1318	-0.0374	-0.1140
50	-0.0655	0.1215	-0.1732	0.1717	-0.0787	-0.0675	0.1416	-0.0542	-0.0970	0.1049
55	-0.0231	0.0541	-0.1036	0.1569	-0.1753	0.1159	0.0151	-0.1262	0.1053	0.0357
60	-0.0068	0.0190	-0.0448	0.0881	-0.1402	0.1715	-0.1404	0.0331	0.0930	-0.1291
65	-0.0017	0.0060	-0.0156	0.0372	-0.0747	0.1239	-0.1628	0.1544	-0.0733	-0.0514
70	-3.9 $\times 10^{-4}$	0.0014	-0.0046	0.0129	-0.0309	0.0633	-0.1086	0.1514	-0.1601	0.1043
75	-8.1 $\times 10^{-5}$	3.3 $\times 10^{-4}$	-0.0012	0.0038	-0.0107	0.0257	-0.0553	0.0946	-0.1386	0.1595
80	-1.6 $\times 10^{-5}$	7.1 $\times 10^{-5}$	-2.9 $\times 10^{-4}$	0.0010	-0.0032	0.0088	-0.0214	0.0453	-0.0821	0.1253
85	-2.8 $\times 10^{-6}$	1.4 $\times 10^{-5}$	-6.2 $\times 10^{-5}$	2.5 $\times 10^{-4}$	-8.6 $\times 10^{-4}$	0.0027	-0.0074	0.0179	-0.0383	0.0710
90	-4.8 $\times 10^{-7}$	2.6 $\times 10^{-6}$	-1.3 $\times 10^{-5}$	5.4 $\times 10^{-5}$	-2.1 $\times 10^{-4}$	7.2 $\times 10^{-4}$	-0.0022	0.0061	-0.0150	0.0323
95	-7.9 $\times 10^{-8}$	4.6 $\times 10^{-7}$	-2.4 $\times 10^{-6}$	1.1 $\times 10^{-5}$	-4.7 $\times 10^{-5}$	1.8 $\times 10^{-4}$	-6.1 $\times 10^{-4}$	0.0019	-0.0051	0.0125
100	-1.2 $\times 10^{-8}$	7.8 $\times 10^{-8}$	-4.4 $\times 10^{-7}$	2.2 $\times 10^{-6}$	-1.0 $\times 10^{-5}$	4.1 $\times 10^{-5}$	-1.5 $\times 10^{-4}$	5.1 $\times 10^{-4}$	-0.0016	0.0043

Table 5
ZEROS OF LAGUERRE FUNCTIONS

n	$\alpha_{n,1}$	$\alpha_{n,2}$	$\alpha_{n,3}$	$\alpha_{n,4}$	$\alpha_{n,5}$	$\alpha_{n,6}$	$\alpha_{n,7}$	$\alpha_{n,8}$	$\alpha_{n,9}$	$\alpha_{n,10}$	$\alpha_{n,11}$	$\alpha_{n,12}$	$\alpha_{n,13}$	$\alpha_{n,14}$	$\alpha_{n,15}$	$\alpha_{n,16}$	$\alpha_{n,17}$	$\alpha_{n,18}$	$\alpha_{n,19}$	$\alpha_{n,20}$
1	0																			
2	0.585 79	3.414 21																		
3	0.415 78	2.294 28	6.289 94																	
4	0.322 55	1.745 76	4.536 62	9.395 07																
5	0.263 56	1.413 41	3.596 43	7.083 81	12.640 79															
6	0.222 85	1.188 94	2.992 74	5.775 14	9.837 46	15.982 87														
7	0.193 043	1.026 67	2.567 88	4.900 35	7.045 91	12.734 18	19.395 72													
8	0.170 279	0.903 71	2.231 09	4.266 70	6.075 36	10.785 36	15.746 93	22.963 13												
9	0.152 317	0.807 23	2.003 14	3.783 47	5.352 30	9.330 15	13.466 93	18.833 59	26.374 07											
10	0.137 789	0.729 46	1.808 35	3.401 43						29.920 70										
11	0.125 792	0.665 42	1.647 16	3.091 13	5.029 29	7.509 89	10.605 95	14.431 61	19.178 86	25.217 70	33.497 20									
12	0.115 718	0.611 76	1.512 62	2.833 75	4.599 23	6.844 53	9.621 31	13.006 05	17.116 85	22.151 09	28.487 96	37.079 13								
13	0.107 139	0.566 14	1.396 58	2.616 60	4.238 85	6.292 26	8.815 00	11.861 40	15.510 76	19.884 63	25.185 25	31.800 38	40.723 01							
14	0.099 744	0.526 87	1.300 64	2.430 81	3.932 11	5.825 53	8.140 24	10.916 50	14.210 81	18.104 89	22.723 37	28.272 97	35.149 44	44.366 08						
15	0.093 305	0.492 70	1.215 60	2.269 96	3.667 63	5.425 33	7.565 93	10.120 22	13.130 29	16.654 41	20.776 47	25.623 88	31.407 51	38.530 68	48.026 08 (5)					
16	0.087 647	0.462 70	1.141 06	2.129 29	3.437 10	5.078 02	7.070 34	9.438 31	12.214 22	15.441 53	19.180 15	23.515 00	28.578 72	34.583 40	41.940 45	51.701 16				
17	0.082 636	0.436 16	1.075 18	2.005 20	3.234 27	4.773 51	6.637 83	8.846 68	11.475 59	14.573 24	17.838 38	21.778 26	26.315 31	31.581 77	37.796 09	45.375 72	55.389 75			
18	0.078 167	0.412 50	1.016 52	1.894 90	3.054 36	4.504 20	6.256 72	8.328 82	10.732 43	13.573 88	16.683 21	20.310 76	24.440 67	29.168 21	34.627 92	41.041 81	48.833 93	59.090 55		
19	0.074 157	0.391 28	0.963 96	1.796 19	2.893 66	4.264 21	5.918 18	7.459 01	9.594 40	12.730 88	15.691 38	19.048 99	22.850 84	27.160 66	32.069 12	37.712 89	44.317 36	52.312 91	62.802 42	
20	0.070 538	0.372 14	0.916 59	1.707 31	2.749 21	4.048 92	5.613 18			12.038 80	14.814 30	17.948 90	21.478 78	25.451 70	29.932 55	35.013 43	40.833 05	47.619 96	55.810 81	66.524 41

In practice, however, we may not know k and the α_s initially. The following procedure is suggested when we know $f(x)$ as a function of x (graphically or otherwise) at a number of points in the interval $(0, x_0)$, and we also know that $|f(x)|$ is negligible for $x > x_0$.

The nature of the particular problem will give us a sufficient idea of the maximum number N of terms of a Laguerre series that is justified by the accuracy of the data. We then choose k so that for $x > x_0$, $r \leq N - 1$, $|\lambda_r(kx)| < 10^{-4}$. The Tables of $\lambda_n(x)$ enable us to determine kx_0 as follows:

N	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
kx_0	100	95	90	85	80	75	70	70	65	60	55	50	45	36	32

Then, by graphical or other interpolation, we determine the values of f when $x = \alpha_{Nr}/k$ ($r = 1, 2, \dots, n$), α_{Nr} being, as in the Table of Zeros, the r th zero of $\lambda_N(x)$ in ascending order. If

$$f(x) = \sum_{s=0}^{N-1} b_s \lambda_s(kx) \dots \dots \dots (19)$$

we then have to solve the simultaneous equations

$$\sum_{s=0}^{N-1} b_s \lambda_s(\alpha_{Nr}) = f(\alpha_{Nr}/k) \dots \dots \dots (20)$$

These equations have an explicit solution

$$b_s = \sum_{r=1}^N \frac{\alpha_{Nr} f(\alpha_{Nr}/k) \lambda_s(\alpha_{Nr})}{N^2 \lambda_{N-1}^2(\alpha_{Nr})} \dots \dots \dots (21)$$

in which the coefficients

$$\alpha_{Nr} \lambda_s(\alpha_{Nr}) / N^2 \lambda_{N-1}^2(\alpha_{Nr}) \dots \dots \dots (22)$$

can be worked out in advance for various values of N , as suggested in Reference 8. The relations between quantities $\lambda'_N(\alpha_{Nr})$, $\lambda_s(\alpha_{Nr})$ are discussed more fully (without proof) in Section 13.2.

(5) DIFFERENTIATION

It can be shown directly from (2) and the recurrence relation (4) for polynomials L_n instead of functions λ_n , that

$$L'_n(x) = - \sum_{s=0}^{n-1} L_s(x) \dots \dots \dots (23a)$$

and $xL'_n(x) = n[L_n(x) - L_{n-1}(x)] = -n \int_0^x L_{n-1}(\xi) d\xi \dots \dots \dots (23b)$

It follows from (23a) that

$$\frac{d}{dx} [e^{\alpha x} L_n(x)] = e^{\alpha x} \left[\alpha L_n(x) - \sum_{s=0}^{n-1} L_s(x) \right] \dots \dots (24)$$

and in the particular case for which $\alpha = -\frac{1}{2}$ we have

$$\lambda'_n(x) = -\frac{1}{2} \lambda_n(x) - \sum_{s=0}^{n-1} \lambda_s(x) \dots \dots \dots (25)$$

The results (23a), (23b), (24) and (25) are valid for all values of x , and are to be distinguished from the results given below in (28), (29) and (30).

The orthonormal properties of $\lambda_n(x)$ are restricted to the range $0 \leq x < \infty$. Particular interest therefore attaches to the discontinuous functions, $L_n(x)H(x)$, $\lambda_n(x)H(x)$, where $H(x)$ is the Heaviside unit function defined as

$$H(x) = \begin{cases} 0; & x < 0 \\ \frac{1}{2}; & x = 0 \\ 1; & x > 0 \end{cases} \dots \dots \dots (26)$$

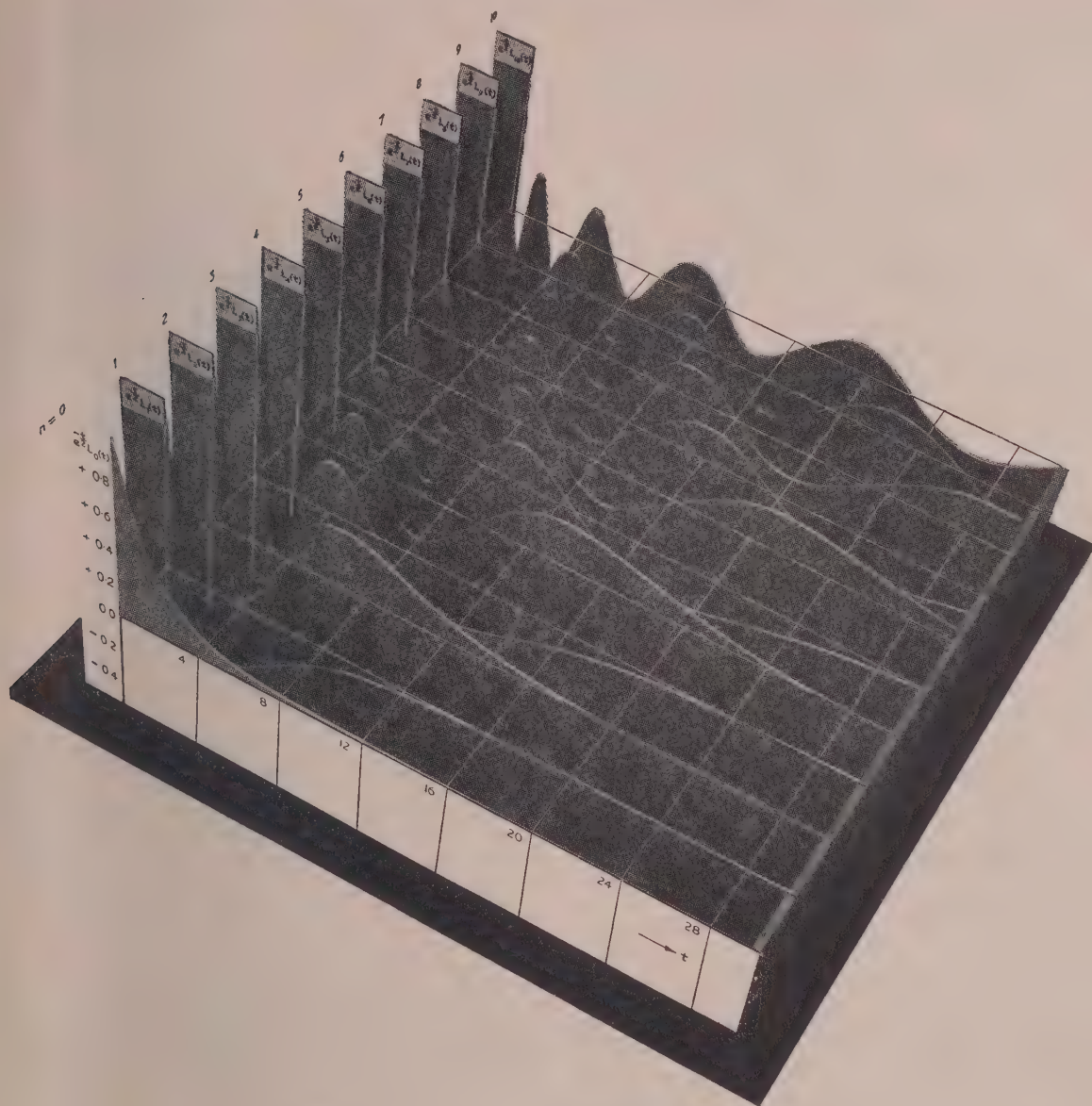


Fig. 1.—March of the Laguerre functions of orders 1 to 10.

In the differentiation of the discontinuous function $L_n(x)H(x)$ we note that $H'(x)$ is the Dirac delta function $\delta(x)$. We obtain, using (23a),

$$\frac{d}{dx}[L_n(x)H(x)] = \delta(x) - H(x) \sum_{s=0}^{n-1} L_s(x) \quad (27)$$

As stated later, $\delta(x)$ [expressions (63) and (68)] may be expanded as an infinite series of Laguerre functions or polynomials. The counterpart of expression (23a) thus becomes

$$\frac{d}{dx}[L_n(x)H(x)] = H(x) \sum_{s=n}^{\infty} L_s(x) \quad (28)$$

in the discontinuous case.

Corresponding to (24), we have

$$\frac{d}{dx}[\varepsilon^{\alpha x} L_n(x)H(x)] = \varepsilon^{\alpha x} H(x) \left[\alpha L_n(x) + \sum_{s=n}^{\infty} L_s(x) \right] \quad (29)$$

and from (29), when $\alpha = -\frac{1}{2}$, we obtain, corresponding to (25),

$$\frac{d}{dx}[\lambda_n(x)H(x)] = H(x) \left[\frac{1}{2}\lambda_n(x) + \sum_{s=n+1}^{\infty} \lambda_s(x) \right] \quad (30)$$

It should be noted that (29) and (30), like (28), imply the presence of a delta function when $x = 0$.

In the four following derivatives, valid for $x \geq 0$, the factor $H(x)$ may be dropped, since the expressions of which the derivative is considered vanish at $x = 0$.

$$\frac{d}{dx}[\lambda_n(x) - \lambda_{n+1}(x)] = \frac{1}{2}[\lambda_n(x) + \lambda_{n+1}(x)] \quad (31)$$

$$\frac{d}{dx}\{\varepsilon^{-1/2x}[\lambda_n(x) - \lambda_{n+1}(x)]\} = \varepsilon^{-1/2x}\lambda_{n+1}(x) \quad (32)$$

$$\frac{d}{dx}\{\varepsilon^{1/2x}[\lambda_n(x) - \lambda_{n+1}(x)]\} = \varepsilon^{1/2x}\lambda_n(x) \quad (33)$$

$$\frac{d}{dx}[x\lambda_n(x)] = \frac{1}{2}[(n+1)\lambda_{n+1}(x) + \lambda_n(x) - n\lambda_{n-1}(x)] \quad (34)$$

The derivative of $\lambda_n(x)$ with regard to the order n can also be found by operational calculus. The result is

$$\frac{\partial}{\partial n}\lambda_n(x) = -\int_0^x \lambda_n(\xi) \frac{\sinh \frac{1}{2}(x-\xi)}{\frac{1}{2}(x-\xi)} d\xi \quad (35a)$$

$$= -\int_0^x \lambda_n(x-\xi) \frac{\sinh \frac{1}{2}\xi}{\frac{1}{2}\xi} d\xi \quad (35b)$$

(6) INTEGRATION

The fundamental equation for integration of Laguerre functions is obtained by reversing (31), namely

$$\int_0^x [\lambda_n(\xi) + \lambda_{n+1}(\xi)] d\xi = 2[\lambda_n(x) - \lambda_{n+1}(x)] \quad (36)$$

Remembering that $\lambda_0(x)$ is $\varepsilon^{-1/2x}$, we deduce, by repeated application of (36),

$$\int_0^x \lambda_n(\xi) d\xi = 2 \left[(-1)^n + 2 \sum_{s=0}^{n-1} (-1)^{n+s+1} \lambda_s(x) - \lambda_n(x) \right] \quad (37)$$

$$\int_0^x [\lambda_n(\xi) - (-1)^r \lambda_{n+r}(\xi)] d\xi = 2 \sum_{s=0}^{r-1} (-1)^s [\lambda_{n+s}(x) - \lambda_{n+s+1}(x)] \quad (38)$$

(32) and (33) can also be reversed to give

$$\int_0^x \varepsilon^{-1/2\xi} \lambda_n(\xi) d\xi = \varepsilon^{-1/2x} [\lambda_{n-1}(x) - \lambda_n(x)] \quad (39)$$

$$\int_0^x \varepsilon^{1/2\xi} \lambda_n(\xi) d\xi = \varepsilon^{1/2x} [\lambda_n(x) - \lambda_{n+1}(x)]^* \quad (40)$$

If we make x tend to infinity in (37), we have

$$\int_0^\infty \lambda_n(\xi) d\xi = 2(-1)^n \quad (41)$$

The orthonormal relation (13) should perhaps be noted again here; in particular if $m = 0$, $n \neq 0$ it gives

$$\int_0^\infty \varepsilon^{-1/2\xi} \lambda_n(\xi) d\xi = 0 \quad (n \neq 0) \quad (42)$$

The following definite integrals can be obtained operationally:

$$\int_0^\infty \varepsilon^{-\alpha x} \lambda_n(\beta x) dx = \frac{(\alpha - \frac{1}{2}\beta)^n}{(\alpha + \frac{1}{2}\beta)^{n+1}} \quad (43)$$

$$\int_0^\infty \varepsilon^{-1/2(\alpha+\beta)x} \lambda_m(\alpha x) \lambda_n(\beta x) dx = \frac{\alpha^n \beta^m}{(\alpha + \beta)^{m+n+1}} \frac{(m+n)!}{m!n!} \quad (44)$$

$$\int_0^\infty [\varepsilon^{-1/2x} x^m \lambda_n(x) dx] / m! = \begin{cases} (-1)^n (m!) / [(m-n)! n!] & m > n \\ (-1)^n & m = n \\ 0 & m < n \end{cases} \quad (45)$$

* The integral in (40) may also be expressed, by means of the convolution theorem of the operational calculus, as

$$\varepsilon^{1/2x} \int_0^x \lambda_{n-m}(x-\xi) \lambda_m(\xi) d\xi; \quad m \text{ real.}$$

$$\left. \begin{aligned} \int_0^\infty \lambda_m[\alpha(t+x)] \lambda_n(\alpha t) dt \\ = \frac{1}{\alpha} \varepsilon^{-1/2\alpha x} [L_{m-n}(\alpha x) - L_{m-n-1}(\alpha x)] \quad (m > n) \\ = \frac{1}{\alpha} \varepsilon^{-1/2\alpha x} \quad (m = n) \\ = 0 \quad (m < n) \end{aligned} \right\} \quad (46)^*$$

(7) RELATIONS BETWEEN LAGUERRE POLYNOMIALS OF ARGUMENT x AND THOSE OF ARGUMENT kx (k REAL AND POSITIVE)

The following results can easily be proved operationally (see Section 9).

$$L_n(kx) = \sum_{s=0}^n \binom{n}{s} k^s (1-k)^{n-s} L_s(x) \quad (47)$$

$$L_n(x) = k^{-n} \sum_{s=0}^n \binom{n}{s} (k-1)^{n-s} L_s(kx) \quad (48)$$

In particular, when $k = 2$

$$L_n(2x) = \sum_{s=0}^n (-1)^{n-s} \binom{n}{s} 2^s L_s(x) \quad (49)$$

$$L_n(x) = 2^{-n} \sum_{s=0}^n \binom{n}{s} L_s(2x) \quad (50)$$

$\binom{n}{s}$ being a binomial coefficient as in (3).

These results can also be expressed in terms of Laguerre functions if an appropriate exponential factor is included on the right-hand side.

A relation also exists between the Laguerre functions of positive order and negative argument, of negative order and positive argument, and the functions $\lambda_n(x)$. This is expressed by

$$\lambda_n(-x) = \lambda_{-(n+1)}(x) = \varepsilon^x \sum_{s=0}^n (-1)^s \binom{n}{s} 2^{n-s} \lambda_s(x) \quad (51)$$

This triple relation enables the present Tables to be extended into the negative range of order and argument.

The Laguerre functions of half-integral order are related to the Bessel functions of the first kind, of imaginary argument. Thus it can be shown, by operational methods, that

$$\lambda_{-1/2}(x) = I_0(\frac{1}{2}x) \quad (52)$$

$$\lambda_{1/2}(x) - \lambda_{-1/2}(x) = x [I_1(\frac{1}{2}x) - I_0(\frac{1}{2}x)] \quad (53)$$

From these two relations we can derive expressions for $\lambda_{(m+1/2)}(x)$, $m = 0, 1, 2, \dots$, in terms of the above Bessel functions, by means of the Laguerre recurrence formula (4).

(8) ADDITION FORMULAE

In a similar manner, we can prove

$$L_n(\alpha + \beta)x = \frac{\alpha^n}{\beta^n} \sum_{s=0}^n (-1)^{n-s} \binom{n}{s} \left(1 + \frac{\beta}{\alpha}\right)^s L_s(\beta x) \quad (54)$$

and since the left-hand side of (54) is symmetrical in α, β we may exchange α, β on the right-hand side to obtain also

$$L_n(\alpha + \beta)x = \frac{\beta^n}{\alpha^n} \sum_{s=0}^n (-1)^{n-s} \binom{n}{s} \left(1 + \frac{\alpha}{\beta}\right)^s L_s(\alpha x) \quad (55)$$

* The authors are indebted to Mr. D. G. Lampard for this formula.

Multiplying (54) by β^n and (55) by α^n and adding, we obtain the alternative symmetrical formula

$$(\alpha^n + \beta^n)L_n[(\alpha + \beta)x] = \sum_{s=0}^n (-1)^s \binom{n}{s} (\alpha + \beta)^{n-s} [\alpha^s L_{n-s}(\beta x) + \beta^s L_{n-s}(\alpha x)]. \quad (56)$$

(9) EXPANSIONS

The expansions set forth in the following Table are all obtained heuristically; typical examples are discussed fully in Sections 13.4 and 13.5; further details are in Reference 9. In essence the method adopted in deriving these expansions is to make use of well-established theorems in modern operational calculus,⁵ and of the three principal operational transforms of the Laguerre polynomials and functions:

$$L_n(x)H(x) = \varepsilon^{1/2x} \lambda_n(x)H(x) \div \left(1 - \frac{1}{p}\right)^n. \quad (57)$$

$$\varepsilon^{-1/2x} L_n(x)H(x) = \lambda_n(x)H(x) \div \frac{p}{p + \frac{1}{2}} \left(\frac{p - \frac{1}{2}}{p + \frac{1}{2}}\right)^n. \quad (58)$$

$$\varepsilon^{-x} L_n(x)H(x) = \varepsilon^{-1/2x} \lambda_n(x)H(x) \div \left(\frac{p}{p + 1}\right)^{n+1}. \quad (59)$$

In (57), (58) and (59), we use Humbert's notation*

$$F[p] \div F(x)$$

as an abbreviation for

$$F[p] = p \int_0^\infty \varepsilon^{-px} F(x) dx \quad (60)$$

$F[p]$ in (60) is the transform of a function $F(x)$ [or rather of $F(x)H(x)$] for the range $x \geq 0$; it is manipulated algebraically to give an expansion of one of the three forms

$$\sum_{s=0}^N A_s \left(1 - \frac{k}{p}\right)^s, \quad \frac{p}{p + \frac{1}{2}k} \sum_{s=0}^N B_s \left(\frac{p - \frac{1}{2}k}{p + \frac{1}{2}k}\right)^s, \quad \sum_{s=0}^N C_s \left(\frac{p}{p + k}\right)^{s+1} \quad (61)$$

where $0 < N \leq \infty$; $0 < k < \infty$

The convergence of the corresponding Laguerre expansions in the x domain when N is infinite is in general assured, provided that the series in the p domain converge over a range of positive real values of p .^{5,10} The rate of convergence is discussed in Section 13.3.

(10) TABLE OF LAGUERRE EXPANSIONS†

(10.1) Elementary Expansions (Range $0 \leq x \leq \infty$)

$$\frac{x^n}{n!} = \sum_{s=0}^n (-1)^s \binom{n}{s} L_s(x) \quad (\text{Laguerre}) \quad (62)$$

$$\delta(x) = \sum_{s=0}^\infty L_s(x) \quad (63)$$

$$H(x) = L_0(x) \quad (64)$$

$$\varepsilon^{1/2x} = 2 \sum_{s=0}^\infty (-1)^s L_s(x) \quad (65)$$

$$\varepsilon^{-kx} = \sum_{s=0}^\infty k^s (k+1)^{-s-1} L_s(x) \quad (66)$$

$$\varepsilon^{-1/2x} \frac{x^n}{n!} = \sum_{s=0}^n (-1)^s \binom{n}{s} \lambda_s(x) \quad (67)$$

$$\delta(x) = \sum_{s=0}^\infty \lambda_s(x) \quad (68)$$

$$H(x) = 2 \sum_{s=0}^\infty [\lambda_{2s}(x) - \lambda_{2s+1}(x)] \quad (69)$$

$$\cos \frac{1}{2}x = \sum_{s=0}^\infty (-1)^s [\lambda_{2s}(x) + \lambda_{2s+1}(x)] \quad (70)$$

$$\sin \frac{1}{2}x = \sum_{s=0}^\infty (-1)^s [\lambda_{2s}(x) - \lambda_{2s+1}(x)] \quad (71)$$

$$\varepsilon^{-kx} = \sum_{s=0}^\infty (k - \frac{1}{2})^s (k + \frac{1}{2})^{-s-1} \lambda_s(x) \quad (72)$$

(10.2) Delayed Functions (Range $\xi < x < \infty$)

$$\delta(x - \xi) = k \sum_{s=0}^\infty \lambda_s(k\xi) \lambda_s(kx) \quad (73)$$

$$H(x - \xi) = k \sum_{s=0}^\infty \lambda_s(k\xi) \int_0^x \lambda_s(ky) dy \quad (74)$$

$$\varepsilon^{-(x-\xi)/2} = \sum_{s=0}^\infty \lambda_s(\xi) [\lambda_s(x) - \lambda_{s+1}(x)] \quad (75)$$

$$\lambda_n(x - \xi) = \sum_{s=0}^\infty \lambda_s(\xi) [\lambda_{n+s}(x) - \lambda_{n+s+1}(x)] \quad (76)$$

(10.3) Circular and Exponential Functions

$$\cos \frac{1}{2}\omega x = 2k(k^2 + \omega^2)^{-1/2} \sum_{s=0}^\infty (-1)^s T_{2s+1}[k/(k^2 + \omega^2)^{1/2}] \lambda_s(kx) \quad (77)$$

$$\sin \frac{1}{2}\omega x = 2k(k^2 + \omega^2)^{-1/2} \sum_{s=0}^\infty (-1)^s U_{2s+1}[k/(k^2 + \omega^2)^{1/2}] \lambda_s(kx) \quad (78)$$

$$\begin{aligned} \varepsilon^{-1/2kx} \cos \omega x &= k(k^2 + \omega^2)^{-1/2} \sum_{s=0}^\infty \omega^s (k^2 + \omega^2)^{-1/2s} U_{s+1}[\omega/(k^2 + \omega^2)^{1/2}] \lambda_s(kx) \\ &\quad (79)^* \end{aligned}$$

$$\begin{aligned} \varepsilon^{-1/2kx} \sin \omega x &= k(k^2 + \omega^2)^{-1/2} \sum_{s=0}^\infty \omega^s (k^2 + \omega^2)^{-1/2s} T_{s+1}[\omega/(k^2 + \omega^2)^{1/2}] \lambda_s(kx) \\ &\quad (80)^* \end{aligned}$$

In (77), (78), (79) and (80) the T 's and U 's are Chebyshev functions, as defined by van der Pol and Weijers,¹¹ namely

$$T_n(\mu) = \cos(n \cos^{-1} \mu); \quad U_n(\mu) = \sin(n \cos^{-1} \mu)$$

(10.4) Bessel Functions of Zero Order

$$J_0[2\sqrt{(\alpha x)}] = \varepsilon^{-\alpha/k} \sum_{s=0}^\infty \frac{1}{s!} \left(\frac{\alpha}{k}\right)^s L_s(kx) \quad (81)$$

$$J_0(x \cot \psi) = \sin \theta \sum_{s=0}^\infty \cos^s \theta P_s(\cos \theta) L_s(kx); \quad (\tan \theta = k \tan \psi) \quad (82)$$

$$J_0(x \cot \psi) = 2 \sin \theta \sum_{s=0}^\infty P_s(\cos 2\theta) \lambda_s(2kx); \quad (\tan \theta = k \tan \psi) \quad (83)$$

* HUMBERT, P.: 'Le calcul symbolique' (Herman et Cie, 1934).

† In Sections 10.2, 10.3 and 10.4 the parameter k , which appears on the right-hand sides of the expansions, is arbitrary, real and positive.

* The authors are indebted to Mr. D. G. Lampard for these two formulae.

$$J_0[\sqrt{(\alpha x)}] = 2 \sum_{s=0}^{\infty} (-1)^s \lambda_s(\alpha k) \lambda_s\left(\frac{x}{k}\right) \quad (84)$$

$$J_0(x) = 2 \sum_{s=0}^{\infty} (-1)^s \lambda_s(kx) \lambda_s\left(\frac{x}{k}\right) \quad (85)$$

$$J_0\left[\frac{2x\sqrt{(k\alpha\beta)}}{1+k}\right] = (1+k) \exp\left[\frac{1}{2}(\alpha+\beta)\frac{1-k}{1+k}\right] \sum_{s=0}^{\infty} (-k)^s \lambda_s(\alpha x) \lambda_s(\beta x) \quad (86)$$

$$J_0(\alpha x) = \frac{2k}{(\alpha^2 + k^2)^{1/2}} \sum_{s=0}^{\infty} P_s\left(\frac{\alpha^2 - k^2}{\alpha^2 + k^2}\right) \lambda_s(2kx) \quad (87)$$

$$J_0\left[x \frac{r(1-k^2)^{1/2}}{r^2 - 2kr + 1}\right] = (r^2 - 2kr + 1)^{1/2} \exp\left(\frac{1}{2}x \frac{1-r^2}{r^2 - 2kr + 1}\right) \sum_{s=0}^{\infty} r^s P_s(k) \lambda_s(x) \quad (88)$$

$$\varepsilon^{-(k-1)x} I_0(x) = \varepsilon^{-(k-1)x} \lambda_{-1/2}(2x) \quad [\text{cf. (52)}]$$

$$= [k/(k-1)\pi]^{1/2} \sum_{s=0}^{\infty} (-1)^s \Gamma(s + \frac{1}{2}) \lambda_s(2kx)/s! \quad (89)$$

In (82), (83), (87) and (88) the P 's are Legendre polynomials.

(10.5) Kelvin Functions of Order Zero

$$\text{bei } \sqrt{(2\pi x)} = \sum_{s=0}^{\infty} (-1)^s \pi^{2s} L_{2s}(x) / [2^{2s}(2s)!] \quad (90)$$

$$\text{bei } 2\sqrt{(\pi x)} = \sum_{s=0}^{\infty} (-1)^s \pi^{2s+1} L_{2s+1}(x) / (2s+1)! \quad (91)$$

$$\text{ber } \sqrt{(2\pi x)} = \sum_{s=0}^{\infty} (-1)^s \pi^{2s+1} L_{2s+1}(x) / 2^{2s+1}(2s+1)! \quad (92)$$

$$-\text{ber } 2\sqrt{(\pi x)} = \sum_{s=0}^{\infty} (-1)^s \pi^{2s} L_{2s}(x) / (2s)! \quad (93)$$

$$\cos \alpha \text{ ber } 2\sqrt{(\alpha x)} + \sin \alpha \text{ bei } 2\sqrt{(\alpha x)} = \sum_{s=0}^{\infty} (-1)^s \alpha^{2s} L_{2s}(x) / (2s)! \quad (94)$$

$$\sin \alpha \text{ ber } 2\sqrt{(\alpha x)} - \cos \alpha \text{ bei } 2\sqrt{(\alpha x)} = \sum_{s=0}^{\infty} (-1)^s \alpha^{2s+1} L_{2s+1}(x) / (2s+1)! \quad (95)$$

(10.6) Incomplete Gamma Function; Error Integral

$$\begin{aligned} \gamma(n+1, x)/n! &= \int_0^x (\varepsilon^{-\xi} \xi^n d\xi)/n! \\ &= \varepsilon^{-x} \sum_{s=0}^{n+1} (-1)^s \binom{n+1}{s} L_{s-1}(x) \quad (96) \\ &= \varepsilon^{-x} \sum_{s=0}^n (-1)^s \binom{n}{s} [L_{s-1}(x) - L_s(x)] \quad (97) \end{aligned}$$

In (96) and (97), $L_{-1}(x)$ is to be interpreted as ε^x , and n is an integer, but if we now put $n = -\frac{1}{2}$ we can extend (96) to obtain a result which can be expressed in the form

$$\sqrt{\pi}/\text{erf}\sqrt{x} = \varepsilon^{-x} \sum_{s=0}^{\infty} \Gamma(\frac{1}{2} + s) L_{s-1}(x)/s! \quad (98)$$

$\gamma(n+1, x)$ is also known as Prym's function $P(x, n+1)$.

(11) ACKNOWLEDGMENTS

The authors' thanks and acknowledgments are due to the Assistant Secretary of the Royal Society for providing tabulated values of $n! \lambda_n(x)$ calculated for the Mathematical Tables Committee by its computer, Miss Sheila M. Burrough. Acknowledgments are also due to the authors and publishers of Reference 1 for some of the values used in the Tables of $\lambda_n(x)$.

The authors wish also to express their thanks to the referees of the paper for many helpful suggestions and corrections. Acknowledgment is also due to the Director of Technical Services of the British Broadcasting Corporation for permission to publish the paper.

(12) REFERENCES

- (1) JAHNKE-EMDE: 'Tables of Higher Functions' (Teubner, 1952).
- (2) LEE, Y. W.: 'Synthesis of Electrical Networks by means of the Fourier Transforms of Laguerre's Functions', *Journal of Mathematics and Physics*, 1931-32, **11**, pp. 83-113.
- (3) WIENER, N.: 'Extrapolation, Interpolation and Smoothing of Stationary Time Series' (Wiley, 1949). (This reference contains tables of six Laguerre functions with Wiener's normalization.)
- (4) WAGNER, K. W.: 'Operatorenrechnung und Laplacesche Transformation' (J. A. Barth Verlag, 1950).
- (5) VAN DER POL, B., and BREMMER, H.: 'Operational Calculus based on the Two-sided Laplace Transform' (Cambridge University Press, 1950).
- (6) LAMPARD, D. G.: 'A New Method of determining Correlation Functions of Stationary Time Series', *Proceedings I.E.E.*, Monograph No. 104, August, 1954 (**102 C**, p. 35).
- (7) LAGUERRE, E. N.: 'Sur l'intégrale $\int_0^{\infty} (e^{-x} dx)/x$ ', *Bulletin de la Société mathématique de France*, 1879.
- (8) HEAD, J. W.: 'The Decomposition of Functions', *Proceedings of the Cambridge Philosophical Society*, 1952, **48**, pp. 742-43.
- (9) WILSON, W. PROCTOR: 'Transient Response of an All-pass Network in Terms of a Finite Series of Laguerre Functions', *B.B.C. Research Department Report No. F-023* (1955/15).
- (10) TRICOMI, F.: 'Trasformazione di Laplace et polinomi di Laguerre', *R.C. Accad. Lincei*, 1935.
- (11) VAN DER POL, B., and WEIJERS, T. J.: 'Tchebycheff Polynomials and their relation to Circular Functions, Bessel Functions and Lissajous Figures', *Physica*, Dec., 1933, **1**, pp. 78-96.
- (12) MORRIS, J., and HEAD, J. W.: 'The "Escalator" Process for the Solution of Lagrangian Frequency Equations', *Philosophical Magazine*, 1944, **35**, pp. 735-59.
- (13) WARD, E. E.: 'The Calculation of Transients in Dynamical Systems', *Proceedings of the Cambridge Philosophical Society*, 1954, **50**, pp. 49-59.
- (14) HEAD, J. W.: 'Approximation to Transients by means of Laguerre Series', *ibid.* (to be published).
- (15) HEAD, J. W.: 'Widening the Applicability of Lin's Iteration Process for determining Quadratic Factors of Polynomials', *Quarterly Journal of Mechanics and Applied Mathematics* (to be published).
- (16) OLVER, F. J. W.: 'The Evaluation of Zeros of High-Degree Polynomials', *Philosophical Transactions of the Royal Society, A*, 1952, **244**, pp. 385-415.

(13) APPENDICES

(13.1) Calculation of Zeros of $\lambda_n(x)$ [or $L_n(x)$]

Dividing through the recurrence relation (4) by $\lambda_n(x)$, we have

$$(n+1)\frac{L_{n+1}(x)}{L_n(x)} = (2n+1-x) - n\frac{L_{n-1}(x)}{L_n(x)} \quad (99)$$

The zeros of $L_n(x)$ are all real, distinct and positive, and are separated by the zeros of $L_{n-1}(x)$, so the last term of (99) can be expressed in partial fractions in the usual way. If α_{nr} is the r th zero of $L_n(x)$ in ascending order,

$$\frac{nL_{n-1}(x)}{L_n(x)} = \sum_{r=1}^n \frac{\alpha_{nr}}{\alpha_{nr} - x} \quad (100)$$

since we can express (100) in the equivalent form

$$nL_{n-1}(\alpha_{nr}) + \alpha_{nr}L'_n(\alpha_{nr}) = 0 \quad (101)$$

which is a special case of (23b) with $x = \alpha_{nr}$.

Substitute from (100) into (99), and transpose the partial fraction involving α_{nr} if we seek the zero $\alpha_{n+1,r}$. We then have

$$\frac{\alpha_{nr}}{\alpha_{nr} - x} = (2n+1-x) - \sum_{s=1}^{n'} \frac{\alpha_{ns}}{\alpha_{ns} - x} \quad (102)$$

where the symbol \sum' denotes that the term for which $s=r$ is omitted, and $\alpha_{n+1,r}$ is the as yet unknown value of x nearest to α_{nr} which satisfies (102). If x_0 is an approximation to $\alpha_{n+1,r}$, we obtain a second approximation x_1 by substituting x_0 for x on the right-hand side of (102). Repeating the process, we obtain a third approximation x_2 by substituting x_1 for x in the right-hand side of (102). If x_0 is sufficiently close to α_{nr} , we find that the differences $d_1 = (x_1 - x_0)$, $d_2 = (x_2 - x_1)$, etc., form a geometric progression whose common ratio ρ is very small when r is small compared to n , but $|\rho|$ may approach and surpass unity when r is nearly equal to n . Instead of repeating the process with x_2 substituted for x on the right-hand side of (102), we substitute

$$(x_1^2 - x_0x_2)/(2x_1 - x_0 - x_2)^* \quad (103)$$

and (103) was usually found to be very close to $\alpha_{n+1,r}$, but might have to be regarded as a new x_0 occasionally. To obtain a good starting approximation x_0 , let $s_{n+1,r}$ denote the right-hand side of (102) when $x = \alpha_{n+1,r}$; this quantity is necessarily found in the process of determining $\alpha_{n+1,r}$. Then if $r \leq \frac{1}{2}n$, we may start with

$$x_0 = \alpha_{nr} - \alpha_{nr}/(s_{n-2,r} - 3s_{n-1,r} + 3s_{n,r}) \quad (104)$$

based on a quadratic rule for $s_{n+q,r}$ ($q = -2, -1, 0$ and 1).

If $r \geq \frac{1}{2}n$, we may start with

$$x_0 = \alpha_{nr} - \alpha_{nr}/(s_{n-2,r-3} - 3s_{n-1,r-2} + 3s_{n,r-1}) \quad (105)$$

based on a quadratic rule for $s_{n+q,r+q-1}$ ($q = -2, -1, 0$ and 1).

It is preferable to calculate the zeros in the order

$$\alpha_{n1}, \alpha_{n2}, \dots, \alpha_{n, n-2}, \alpha_{nn} \quad (106)$$

in this way, leaving $\alpha_{n, n-1}$ to be determined from the fact that the sum of the α 's for a given n is n^2 . This is a very sensitive check on numerical working, as if there is an error in the calculation of the zeros (106), a wrong value of $\alpha_{n, n-1}$ will be obtained which will not satisfy (102). An additional, but less sensitive, check is that the product of the α 's of a given order is $n!$.

* This formula is due to A. C. Aitken.

(13.2) Orthogonal and Other Properties of $\lambda_s(\alpha_{nr})$

Consider the recurrence relation (4) between successive Laguerre functions, and the associated set of n homogeneous symmetrical linear simultaneous equations

$$\left. \begin{aligned} -(1-x)\xi_0 + \xi_1 &= 0 \\ \xi_0 - (3-x)\xi_1 + 2\xi_2 &= 0 \\ &\vdots \\ (n-2)\xi_{n-3} - (2n-3-x)\xi_{n-2} + (n-1)\xi_{n-1} &= 0 \\ (n-1)\xi_{n-2} - (2n-1-x)\xi_{n-1} &= 0 \end{aligned} \right\} \quad (107)$$

It is clear that (107) have a non-zero solution

$$\xi_0 : \xi_1 : \xi_2 : \dots : \xi_{n-2} : \xi_{n-1} = \lambda_0(\alpha_{nr}) : \lambda_1(\alpha_{nr}) : \dots : \lambda_{n-2}(\alpha_{nr}) : \lambda_{n-1}(\alpha_{nr}) \quad (108)$$

if $x = \alpha_{nr}$, because $\lambda_n(\alpha_{nr})$ is zero. Moreover, this solution is clearly a 'proper vector', having the 'proper value' α_{nr} , of equations (107) with $x = 0$.

Now suppose that

$$D_r^2 = \lambda_0^2(\alpha_{nr}) + \lambda_1^2(\alpha_{nr}) + \dots + \lambda_{n-1}^2(\alpha_{nr}) \quad (109)$$

Then it is shown in Reference 12 that the quantities $\lambda_s(\alpha_{nr})/D_r$, being the quantities described in that reference as 'rectified modes', have the orthogonal properties

$$\left. \begin{aligned} \sum_{i=0}^{n-1} \lambda_i(\alpha_{nr})\lambda_i(\alpha_{ns}) &= 0 \quad \text{if } r \neq s \\ \sum_{r=1}^n \lambda_i^2(\alpha_{nr})/D_r^2 &= 1, \quad \sum_{r=1}^n \lambda_i(\alpha_{nr})\lambda_j(\alpha_{nr})/D_r^2 = 0 \quad (i \neq j) \end{aligned} \right\} \quad (110)$$

and multiplying the i th equation of (107) [with ξ_s replaced by $\lambda_s(x)$ and the last zero replaced by $-n\lambda_n(x)$] by $\lambda_i(\alpha_{nr})/D_r$ and adding, we can deduce

$$\left. \begin{aligned} (\alpha_{nr} - x) \sum_{s=0}^{n-1} \lambda_s(\alpha_{nr})\lambda_s(x) &= n\lambda_n(x)\lambda_{n-1}(\alpha_{nr}) \\ \lambda_s(x) &= - \sum_{r=1}^n \frac{n\lambda_n(x)\lambda_s(\alpha_{nr})\lambda_{n-1}(\alpha_{nr})}{D_r^2(x - \alpha_{nr})} \end{aligned} \right\} \quad (111)$$

But we also have, by the ordinary rule of partial fractions,

$$\frac{\lambda_s(x)}{\lambda_n(x)} = \sum_{r=1}^n \frac{\lambda_s(\alpha_{nr})}{\lambda_n(\alpha_{nr})(x - \alpha_{nr})} \quad (112)$$

from which it follows that

$$D_r^2 = -n\lambda'_n(\alpha_{nr})\lambda_{n-1}(\alpha_{nr}) \quad (113)$$

In Section 13.1 it was also stated [see (101)] that

$$\lambda'_n(\alpha_{nr}) = -n\lambda_{n-1}(\alpha_{nr})/\alpha_{nr} \quad (114)$$

so that

$$D_r = n|\lambda_{n-1}(\alpha_{nr})|/\alpha_{nr}^{1/2} \quad (115)$$

$$\text{and} \quad \frac{\lambda_s(x)}{\lambda_n(x)} = - \sum_{r=1}^n \frac{\alpha_{nr}\lambda_s(\alpha_{nr})}{n\lambda_{n-1}(\alpha_{nr})(x - \alpha_{nr})} \quad (116)$$

Now since

$$\frac{\lambda_n(x)}{\lambda_n(x)} = 1, \quad \frac{\lambda_{n-1}(x)}{\lambda_n(x)} = - \sum_{r=1}^n \frac{\alpha_{nr}}{n(x - \alpha_{nr})} \quad (117)$$

we deduce from the recurrence relation (4) that

$$\frac{\lambda_{n-2}(x)}{\lambda_n(x)} = - \sum_{r=1}^n \frac{(2n-1-\alpha_{nr})\alpha_{nr}}{n(n-1)(x - \alpha_{nr})} \quad (118)$$

and further relations of this nature are deducible similarly.

(13.3) Expansion of a Finite Series of Exponentials in Laguerre Functions

A typical transient motion of an electrical or dynamical system (with lumped elements) is of the form

$$\phi(x) = \sum_{r=1}^n A_r e^{-\alpha_r x} \quad (119)$$

where the α_r have positive real parts if the system is stable, and the A_r are constants, which may be complex.

A Laguerre series approximation to (119) can be obtained by using (72) for each term of the right-hand side; (72) is valid for k complex with positive real part. For a stable system, the coefficients obtained when each term $A_r e^{-\alpha_r x}$ is expressed as a Laguerre series with argument $2bx$ will decrease in geometric progression with common ratio $(\alpha_r - b)/(\alpha_r + b)$ if the α_r are distinct, and we are at liberty to choose b so that the convergence shall be as rapid as possible. Ward¹³ suggests that b should be the geometric mean of the α_r (which is easily determined from the characteristic equation whose roots are the α_r) unless a more detailed knowledge of the α_r and A_r suggests that b should mainly be chosen to obtain rapid convergence for the terms associated with large $|A_r|$. If the α_r are not distinct, but there is an m -fold root α , the corresponding term in $\phi(x)$ will have A_r replaced by a polynomial in x of degree $(m-1)$, and (72) will be replaced by an expression of similar form except that the coefficient of $\lambda_s(bx)$ will now contain a linear combination of the coefficients of α^s in the binomial expansions of $(1+\alpha)^{-v}$ for $v = 1, 2, \dots, m$. The ultimate convergence of the Laguerre series will thus not be affected by the presence of repeated α_r , but for a large value of m more terms may be needed before the trend towards convergence is established. The method described by Ward involves very slow convergence in cases where large A_r are associated with α_r whose real parts are of very different orders of magnitude. In such exceptional cases we suggest that the contribution of large A_r terms to (119) be determined in exponential form and subtracted from $\phi(x)$ before a Laguerre series is found.

If all the α_r are known and the case is not exceptional as described above, the most rapid convergence is obtained¹⁴ for $|A_r|$ of the same order of magnitude, by plotting the points α_r

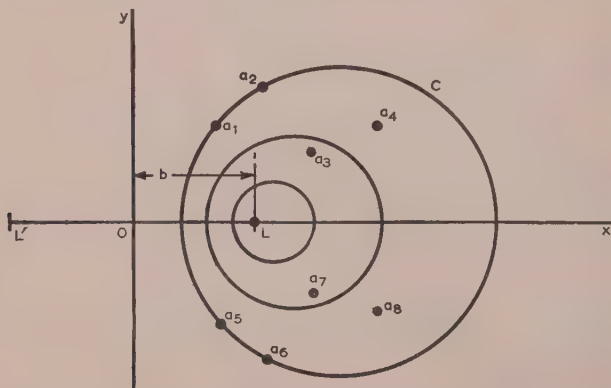


Fig. 2.—The smallest circle C enclosing all the points a_r , and the coaxial system determined by it.

on an Argand diagram (as in Fig. 2) and determining the smallest circle C which contains them all and has its centre on the real axis. If $(\pm b, 0)$ are the limiting points (L, L' in Fig. 2) of the coaxial system with the imaginary axis as radical axis which

includes C, then the Laguerre series should involve functions of argument $2bx$.

The values of α_r most likely to retard convergence of Laguerre series are just those most easily found by Head's adaptation¹⁵ of Lin's iterative method of solving algebraic equations. If the equation whose roots are the α_r is written in the form

$$X[p] \equiv (p^2 + \beta_1)(p^2 + \beta_2) \dots (p^2 + \beta_s) + \kappa p(p^2 + \gamma_1)(p^2 + \gamma_2) \dots (p^2 + \gamma_x) = 0 \quad (120)$$

then for a stable system the β 's and γ 's are real and positive and separate each other, and they are therefore relatively easy to determine.

A useful starting approximation $(p + \alpha_0)$ to a factor of the left-hand side of (120) can often be obtained as follows:

For small κ , let $(p + \alpha_0)$ be one of the factors of

$$(p^2 + \beta_r) + \kappa p \frac{(\gamma_1 - \beta_r)(\gamma_2 - \beta_r) \dots (\gamma_x - \beta_r)}{(\beta_1 - \beta_r)(\beta_2 - \beta_r) \dots (\beta_s - \beta_r)} \quad (121)$$

and for large κ , let $(p + \alpha_0)$ be one of the factors of

$$p^2 + \gamma_r - \frac{p}{\kappa \gamma_r} \frac{(\beta_1 - \gamma_r)(\beta_2 - \gamma_r) \dots (\beta_s - \gamma_r)}{(\gamma_1 - \gamma_r)(\gamma_2 - \gamma_r) \dots (\gamma_x - \gamma_r)} \quad (122)$$

it being understood that zero factors are omitted from the denominators of (121) and (122).

Now suppose that $p + \alpha_0$ is an approximation to a linear factor of $X[p]$. Divide $(p + \alpha_0)$ into X , stopping at the linear term and obtaining a remainder $A_0 p + B_0$. If $B_0/A_0 = \alpha_1$, we now repeat the process with divisor $(p + \alpha_1)$, and obtain remainder $A_1 p + B_1$. If $\alpha_2 = B_1/A_1$, the next divisor of Lin's original process would be $p + \alpha_2$, but we take advantage of the fact that in general $\alpha_1 - \alpha_0$ and $\alpha_2 - \alpha_1$ and similar quantities formed by repetition of the process are nearly in geometric progression; the next divisor is [cf. eqn. (103)]

$$p + [(\alpha_0^2 - \alpha_1 \alpha_2)/(2\alpha_1 - \alpha_0 - \alpha_2)]$$

This process fails for repeated or 'clustered' roots, whose presence can often be detected with sufficient accuracy for determining the circle C in Fig. 2 by performing the 'h.c.f. process' for $X[p]$ and $X'[p]$. If a divisor of the 'h.c.f. process' has abnormally small coefficients, the factors of the divisor before are nearly repeated factors of $X[p]$. The accurate determination of 'clustered' factors of $X[p]$ has been discussed by Olver¹⁶ and will not be further considered here, since it is frequently isolated roots of extreme modulus that retard the convergence of Laguerre series.

Wherever division by $(p + \alpha)$ has been mentioned above with α complex, the simplest procedure is to divide by the real quadratic factor $(p + \alpha)(p + \bar{\alpha})$, $\bar{\alpha}$ being the conjugate of α , until there is a quadratic remainder, and finally to divide the quadratic remainder by $p + \alpha$ to obtain the required linear remainder.

(13.4) Derivation of Laguerre Expansions

The following examples show the derivation of Laguerre expansions for functions

$$e^{-\alpha x} H(x), \quad H(x - \xi), \quad \delta(x - \xi) \quad (123)$$

The operational transform of $e^{-\alpha x} H(x)$ is given by the known relation

$$e^{-\alpha x} H(x) \doteq \frac{p}{p + \alpha} \quad (124)$$

Now

$$\frac{p}{p+\alpha} = \frac{p}{(\alpha + \frac{1}{2})(p + \frac{1}{2}) - (\alpha - \frac{1}{2})(p - \frac{1}{2})}$$

$$= \frac{p}{(\alpha + \frac{1}{2})(p + \frac{1}{2})} \left(1 - \frac{\alpha - \frac{1}{2}}{\alpha + \frac{1}{2}} \frac{p - \frac{1}{2}}{p + \frac{1}{2}} \right)^{-1} \quad (125)$$

$$= \frac{p}{(\alpha + \frac{1}{2})(p + \frac{1}{2})} \sum_{s=0}^{\infty} \left(\frac{\alpha - \frac{1}{2}}{\alpha + \frac{1}{2}} \frac{p - \frac{1}{2}}{p + \frac{1}{2}} \right)^s; \quad [\operatorname{Re}(p) > 0] \quad (126)$$

Since

$$\frac{p}{p + \frac{1}{2}} \left(\frac{p - \frac{1}{2}}{p + \frac{1}{2}} \right)^s \doteq \lambda_s(x) \quad (127)$$

it follows that

$$e^{-\alpha x} H(x) = \frac{1}{\alpha + \frac{1}{2}} \sum_{s=0}^{\infty} \left(\frac{\alpha - \frac{1}{2}}{\alpha + \frac{1}{2}} \right)^s \lambda_s(x) \quad (128)$$

Since the expansion in (128) is an identity, we may substitute p for α , and ξ for x . We then obtain

$$e^{-p\xi} = \frac{1}{p + \frac{1}{2}} \sum_{s=0}^{\infty} \left(\frac{p - \frac{1}{2}}{p + \frac{1}{2}} \right)^s \lambda_s(\xi); \quad 0 < \xi < \infty \quad (129)$$

We recognize $e^{-p\xi}$ as the operational transform of the 'delayed unit function', i.e.

$$H(x - \xi) \doteq e^{-p\xi} \quad (130)$$

Noting that

$$\frac{1}{p + \frac{1}{2}} \left(\frac{p - \frac{1}{2}}{p + \frac{1}{2}} \right)^s \doteq \int_0^x \lambda_s(y) dy \quad (131)$$

we obtain
$$H(x - \xi) = \sum_{s=0}^{\infty} \lambda_s(\xi) \int_0^x \lambda_s(y) dy \quad (132)$$

By differentiating (132) with regard to x , we further obtain

$$\delta(x - \xi) = \sum_{s=0}^{\infty} \lambda_s(\xi) \lambda_s(x) \quad (133)$$

(13.5) Heuristic Derivation of a Bessel Generating Function for a Laguerre Expansion in Two Variables

We consider the Laguerre series

$$G(\alpha, \beta; x) = e^{-(\alpha+\beta)x} \sum_{s=0}^{\infty} (-k)^s L_s(\alpha x) L_s(\beta x) \quad (134)$$

and suppose that x is a parameter while α, β are operationally related to two new variables p, q by the definition integral⁵ [analogous to (60)]

$$G[[p, q; x]] = pq \int_0^{\infty} \int_0^{\infty} e^{-p\alpha - q\beta} G(\alpha, \beta; x) d\alpha d\beta \quad (135)$$

We can regard (135) as the result of performing two operational transformations analogous to that defined by (60). In one of these transformations α is transformed operationally into p and x, q and β are parameters, while in the other β is transformed into q and x, p and α are regarded as parameters. Provided that (135) exists, these transformations may be performed independently, and the known operational transforms (57), (58) and (59) can be applied (with α for x , or with β for x and q for p). We thus obtain

$$G[[p, q; x]] = \sum_{s=0}^{\infty} (-k)^s \left(\frac{p}{p+x} \right)^{s+1} \left(\frac{q}{q+x} \right)^{s+1} \quad (136)$$

From (136), by the purely algebraic process of summing a geometric progression, we obtain

$$G[[p, q; x]] = \frac{pq}{pq(1+k) + (p+q)x + x^2} \quad (137)$$

$$= \frac{1}{1+k} \frac{q}{q + \frac{x}{1+k}} \frac{p}{p + \frac{x}{1+k}} \left[1 + \frac{\frac{kx}{1+k}}{q + \frac{x}{1+k}} \right] \quad (138)$$

Now let

$$e^{x\beta/(1+k)} G(\alpha, \beta; x) = K(\alpha, \beta; x) \quad (139)$$

and suppose that if $K(\alpha, \beta; x)$ is substituted for $G(\alpha, \beta; x)$ on the right-hand side of (135), the left-hand side becomes $K[[p, q; x]]$. Then, applying the exponential shift theorem to (138), we have

$$K[[p, q; x]] = \frac{1}{1+k} \frac{p}{p + \frac{x}{1+k}} \left[1 + \frac{kx}{(1+k)q} \right] \quad (140)$$

Now let $K[\alpha, q; x]$ be the function operationally equivalent to $K[[p, q; x]]$ when α is transformed into p , and q, x are regarded as parameters; $K[\alpha, q; x]$ is (in this case) also operationally equivalent to $K(\alpha, \beta; x)$ when q is transformed into β , and α, x are regarded as parameters. We thus obtain

$$K[\alpha, q; x] = \frac{1}{1+k} \exp \left\{ -\frac{x\alpha}{1+k} \left[1 + \frac{kx}{(1+k)q} \right] \right\} \quad (141)$$

But since we have the known relation

$$J_0[2\sqrt{(r/q)}] \doteq e^{-(r/q)} \quad (142)$$

(141) becomes

$$K(\alpha, \beta; x) = \frac{1}{1+k} e^{-x\alpha/(1+k)} J_0 \left[\frac{2x\sqrt{(k\alpha\beta)}}{1+k} \right] \quad (143)$$

By purely algebraic manipulation of (143), and substituting for $K(\alpha, \beta; x)$ from (139) and for $G(\alpha, \beta; x)$ from (134), we find

$$J_0 \left[\frac{2x\sqrt{(k\alpha\beta)}}{1+k} \right] = (1+k) e^{-k(\alpha+\beta)x/(1+k)} \sum_{s=0}^{\infty} (-k)^s L_s(\alpha x) L_s(\beta x) \quad (144)$$

If $k = 1$, we deduce

$$J_0[x\sqrt{(\alpha\beta)}] = 2 \sum_{s=0}^{\infty} (-1)^s \lambda_s(\alpha x) \lambda_s(\beta x) \quad (145)$$

and if in (144) we substitute $(-k)$ for k ,

$$I_0 \left[\frac{2x\sqrt{(k\alpha\beta)}}{1-k} \right] = (1-k) e^{k(\alpha+\beta)x/(1-k)} \sum_{s=0}^{\infty} k^s L_s(\alpha x) L_s(\beta x) \quad (146)$$

(13.6) Time Response of Repeated Bilinear Networks expressed in Laguerre Expansions: Preferred Excitation

Certain classes of repeated passive networks* have a transfer function of the form

$$\left(\frac{ap+b}{cp+d} \right)^n \quad (147)$$

where a, b, c, d are real ($c, d \neq 0$), and n is the number of sections. The transfer function of a single section is a bilinear function of p .

* Usually lattice (including all-pass) networks.

This expression may be expanded in the alternative forms

$$\left(\frac{ap+b}{cp+d}\right)^n = \left\{ \left(\frac{b}{d}\right)^n \sum_{s=0}^n \binom{n}{s} \left(\frac{ad}{bc} - 1\right)^s \left(\frac{p}{p+\frac{d}{c}}\right)^s \dots \dots \dots (148a) \right.$$

$$\left. \frac{1}{2^n} \left(\frac{bc+ad}{cd}\right)^n \sum_{s=0}^n (-1)^s \binom{n}{s} \left(\frac{bc-ad}{bc+ad}\right)^s \left[\frac{p-\frac{d}{c}}{p+\frac{d}{c}}\right]^s \dots \dots \dots (148b) \right.$$

To the expansion (148a) corresponds [from (59)] the time series

$$\exp - \left(\frac{d}{2c}t\right) \left(\frac{b}{d}\right)^n \sum_{s=0}^n \binom{n}{s} \left(\frac{ad}{bc} - 1\right)^s \lambda_{s-1} \left(\frac{d}{c}t\right) H(t)^* \quad (149)$$

which is the time response† of the repeated network to impressed excitation $H(t)$.

Now suppose that we postulate, instead of $H(t)$, an exciting function of the operational form

$$\left(\frac{p}{p+\frac{d}{c}}\right)^r \doteq \exp\left(-\frac{d}{2c}t\right) \lambda_{r-1} \left(\frac{d}{c}t\right) H(t) \quad r=0, 1, 2, \dots \quad (150)$$

The response to this excitation, associated with the transfer function expanded in (148a), is then given by

$$\begin{aligned} & \left(\frac{b}{d}\right)^n \sum_{s=0}^n \binom{n}{s} \left(\frac{ad}{bc} - 1\right)^s \left(\frac{p}{p+\frac{d}{c}}\right)^{r+s} \\ & \doteq \exp\left(-\frac{d}{2c}t\right) \left(\frac{b}{d}\right)^n \sum_{s=0}^n \binom{n}{s} \left(\frac{ad}{bc} - 1\right)^s \lambda_{r+s-1} \left(\frac{d}{c}t\right) H(t) \end{aligned} \quad (151)$$

* In this summation $\lambda_{-1} \left(\frac{d}{c}t\right)$ is to be interpreted [from (51)] as $\exp\left(\frac{d}{2c}t\right)$.

† Here we relate a function of p , $F[p]$ to a function of t , $F(t)$ by means of (60), with x replaced by t .

Alternatively consider an exciting function, applied to (148b), of the form

$$\frac{p}{p+\frac{d}{c}} \left[\frac{p-\frac{d}{c}}{p+\frac{d}{c}} \right]^r \doteq \lambda_r \left(2\frac{d}{c}t\right) H(t) \quad (152)$$

The time response is obtained in this case by interpreting the product of the left-hand side of (152) and (148b); it is

$$\frac{1}{2^n} \left(\frac{bc+ad}{cd}\right)^n \sum_{s=0}^n (-1)^s \binom{n}{s} \left(\frac{bc-ad}{bc+ad}\right)^s \lambda_{r+s} \left(2\frac{d}{c}t\right) H(t) \quad (153)$$

It will be observed, in both these cases, that the excitations (150) and (152), which involve Laguerre functions of order r , lead to time responses (151) and (153) which are finite Laguerre series. If r is changed to $(r+k)$, all the Laguerre functions involved in (151) and (153) have their orders increased by k , and the series are otherwise unchanged. This suggests that the excitations (150) and (152) are natural to the network; it is therefore appropriate to call them 'preferred excitations'. The simplest preferred excitations in these two cases are those for which $r=0$. For (148a), from (150) this corresponds to

$$H(t) \equiv \exp\left(-\frac{d}{2c}t\right) \lambda_{-1} \left(\frac{d}{c}t\right) H(t)$$

and for (148b), from (152), it is

$$\exp\left(-\frac{d}{c}t\right) H(t) \equiv \lambda_0 \left(2\frac{d}{c}t\right).$$

If an arbitrary exciting function $F(t)H(t)$ can be expanded as a finite or infinite series of the preferred functions, the time response to this excitation can be obtained immediately. Thus for (148a) we need to expand $F(t)H(t)$ in the form

$$F(t)H(t) = \exp\left(-\frac{d}{2c}t\right) \sum_{j=0}^M A_j \lambda_j \left(\frac{d}{c}t\right) \quad (154)$$

while for (148b) we need to expand $F(t)H(t)$ in the Laguerre function series

$$F(t)H(t) = \sum_{k=0}^N B_k \lambda_k \left(2\frac{d}{c}t\right) \quad (155)$$

and the corresponding time response is also a series of Laguerre functions. M and N may be finite or infinite.

MAGNETIC ENERGY AND ELECTRON INERTIA IN A SUPERCONDUCTING SPHERE

By Professor E. G. CULLWICK, O.B.E., M.A., D.Sc., F.R.S.E., Member.

(The paper was first received 30th January, and in revised form 16th March, 1956. It was published as an INSTITUTION MONOGRAPH in June, 1956.)

SUMMARY

The hypothesis that the magnetic energy of a current circuit is the kinetic energy of the effective conduction electrons, developed in a previous Monograph,¹ is applied to the case of a conducting sphere without resistivity in a uniform magnetic field. A surface current is induced which prevents the growth of a magnetic field within the sphere, and expressions are found for the number and velocity of effective conduction electrons which carry the current. It is found that these electrons are in stable radial equilibrium, moving in circular orbits under the action of magnetic forces.

The well-known Meissner effect in pure superconductors is shown to be an expected rather than an unexpected phenomenon, since its absence would require, under certain conditions, a supercurrent lacking equilibrium.

The theory is shown to lead, by means of a simple assumption, to the basic equations of the London theory of superconductivity, but with a different interpretation of the velocity parameter. Finally, the inertial supercurrent and magnetic field which should be produced by the steady rotation of a superconducting sphere, as deduced by the new theory, are shown to be exactly the same as those forecast by the London theory.

LIST OF SYMBOLS

(Rationalized M.K.S. units)

- A , A = Vector potential, webers/m.
 B , B = Magnetic flux density, webers/m².
 C = A constant.
 e = Electronic charge (a negative quantity), coulombs.
 E , E = Electric field intensity, volts/m.
 f , F = Force, newtons.
 H , H = Magnetic field intensity, AT/m.
 I = Electric current, amp.
 J , J = Electric current density, amp/m².
 j_s = Electric current surface density, amp/m.
 L = Self-inductance, henrys.
 L_m = Mutual inductance, henrys.
 l = Length of circuit, m.
 m = Effective mass of conduction electron, kg.
 M = Mass, kg.
 \mathcal{M} = Magnetic moment, newton-m²/amp.
 n = Number of effective conduction electrons in surface current on sphere, or the number of superconducting electrons per unit volume (London theory).
 p = Perpendicular distance of point on surface of sphere, or in surface layer of current, from the axis, m.
 p_s = Momentum of superconducting electrons, London theory.
 R = Radius of sphere, m.
 r = Radial distance of point from centre of sphere, m.
 r = Resistivity, ohm-m.
 t = Time, sec.
 W , w = Energy, joules.
 v , v = Velocity of effective conduction electrons, m/sec.

v_s , v_s = Velocity of superconducting electrons, London theory, m/sec.

θ = Angle.

μ_0 = Primary magnetic constant, 1.257×10^{-6} , weber/amp-m.

ρ = Linear density of effective conduction charge in filamentary circuit, coulombs/m.

ρ_s = Surface density of effective conduction charge on sphere, coulombs/m².

Φ = Linking magnetic flux, webers.

ϕ = Electrostatic potential, volts.

ψ = Multi-valued scalar potential function, webers.

Λ = Constant in London theory, m/ne^2 .

χ = Superpotential, London theory.

(1) INTRODUCTION

In a previous Monograph¹ the consequences of the hypothesis that the magnetic energy is the kinetic energy of the effective conduction electrons, in a current circuit, were developed. It was shown that the hypothesis enables electron-inertia effects in closed conducting circuits to be included in electromagnetic theory, and that the inconsistency of conventional theory with these effects is due to the original view of Maxwell that the carriers of an electric current in a conductor do not possess inertia. It was suggested that the hypothesis might lead to a consistent macroscopic theory of superconductors.

If a steady or quasi-steady conduction current is due to the motion of electrons of effective mass m , charge e , and velocity v , the basic relation of the hypothesis is

$$A = \frac{m}{e} v \quad \dots \quad (1)^1$$

where A is the vector potential of the current. It follows that if the current changes the self-induced electric field in the conductor is

$$E = -\frac{\partial A}{\partial t} = -\frac{m}{e} \frac{\partial v}{\partial t} \quad \dots \quad (2)^1$$

so that the self-induced e.m.f. may be regarded as the equivalent of the inertial effect of the current.

If the circuit is such that v has the same value everywhere, both across the section of the conductor (which must be considered to be filamentary) and around the circuit, the effective conduction charge per unit length of wire is given by

$$\rho = \frac{m}{e} \frac{I}{L} \quad \dots \quad (3)^1$$

The velocity of the effective conduction electrons is therefore

$$v = \frac{I}{\rho} = \frac{e}{m} \frac{LI}{L} \quad \dots \quad (4)^1$$

According to these relations the effective conduction charge in a circuit is determined by the geometry of the circuit, and the current is carried by a comparatively small number of effective electrons moving with a high velocity.

Correspondence on Monographs is invited for consideration with a view to publication.
 Professor Cullwick is Watson-Watt Professor of Electrical Engineering, St. Andrews University (Queen's College, Dundee).

It should be noted that eqn. (1) is independent of the conduction current density, which for the steady or quasi-steady state is given by $\mathbf{J} = \text{curl } \mathbf{H}$. The equation specifies the available velocity or momentum level which may or may not be occupied by effective current-carrying electrons.

(2) MAGNETIC ENERGY OF COUPLED CIRCUITS

The magnetic energy of two circuits of mutual inductance L_m is

$$W_m = \frac{1}{2}L_1I_1^2 \pm L_mI_1I_2 + \frac{1}{2}L_2I_2^2 \quad . \quad . \quad (5)$$

and it is clear that the mutual magnetic energy cannot be included in the individual kinetic energies of the two currents. Thus, if we express W_m as a function of masses and velocities, of the form

$$W_m = \frac{1}{2}M_1v_1^2 \pm M_mv_1v_2 + \frac{1}{2}M_2v_2^2 \quad . \quad . \quad (6)$$

the coefficient M_m is to be regarded as mutual mass or a product of inertia and the middle term as mutual kinetic energy. The values of M_1 and M_2 are the same as when the circuits are isolated, and $\frac{1}{2}M_1v_1^2 = \frac{1}{2}L_1I_1^2$ is the kinetic energy of the current I_1 acting alone, or the magnetic energy of the component of magnetic field due to I_1 , considered by itself. Thus ρ and v in each circuit are independent of the coupling. If, however, the two circuits are joined in series to become a single circuit, then M_1 , $\pm 2M_m$ and M_2 merge together and ρ and v have values appropriate to the new circuit. A current circuit, considered as a single connected dynamical system, must thus be taken as a whole when its properties are being determined.

(3) A CONDUCTING SPHERE WITHOUT RESISTIVITY IN A UNIFORM MAGNETIC FIELD

(3.1) Induced Surface Current

Consider a solid conducting sphere of zero resistivity placed in a uniform magnetic field of flux density B_0 . Since infinite current density would entail infinite energy, the basic law of conduction $\mathbf{E} = \mathbf{J}r$, when $r = 0$, gives $\mathbf{E} = 0$; and since $\text{curl } \mathbf{E} = -\partial \mathbf{B}/\partial t$ it follows that

$$\mathbf{B} = \text{a constant} \quad . \quad . \quad . \quad (7)$$

inside such a conductor. So if the conductor is originally remote from external sources, with no current in it, the internal magnetic field must always remain zero whatever external field may be applied. As Maxwell² pointed out, such a conductor may be regarded as being impervious to magnetic flux.

When the field B_0 is applied to the sphere, therefore, a current must be induced which will cause a component of flux density, B_s , equal and opposite to B_0 inside the sphere, so that the resultant field remains zero. This current must be limited to the surface, since in the interior $\mu_0 \mathbf{J} = \text{curl } \mathbf{B} = 0$. Hence

$$\mathbf{B}_s = -\mathbf{B}_0 \quad . \quad . \quad . \quad (8)$$

inside the sphere. Outside the sphere the resultant field is distorted, as shown in Fig. 1.

Now in order to produce the component of field B_s inside the sphere, the induced current must be equivalent to a magnetic moment per unit volume equal to $\frac{3}{2}B_s$; so to find the necessary surface current consider a thin slice of the sphere (Fig. 2) perpendicular to B_0 , of radius $p = R \sin \theta$ and thickness $R \sin \theta \delta\theta = p\delta\theta$. The width of the surface of the slice is $R\delta\theta$. If the surface current density is j_s , the current around the periphery of the slice is $j_s R\delta\theta$ and its magnetic moment, $\mu_0(\text{current})(\text{area})$, is

$$\mu_0 j_s \pi R p^2 \delta\theta \quad . \quad . \quad . \quad (9)$$

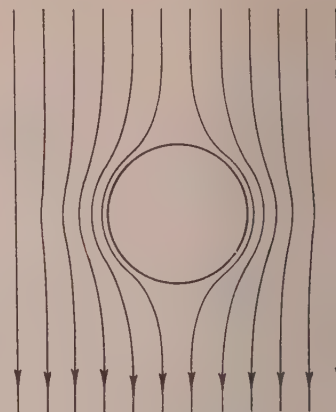


Fig. 1.—Sphere of zero resistivity in magnetic field.

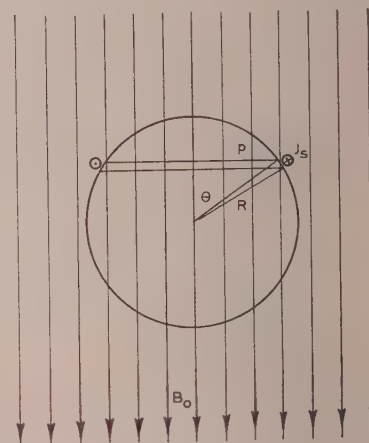


Fig. 2.—Surface current on sphere of zero resistivity in magnetic field.

The volume of the slice is $\pi p^3 \delta\theta$, so that the magnetic moment per unit volume is

$$\mathcal{M} = \mu_0 j_s \frac{R}{p} = \frac{\mu_0 j_s}{\sin \theta} \quad . \quad . \quad . \quad (10)$$

We therefore have $\frac{\mu_0 j_s}{\sin \theta} = \frac{3}{2} B_s$

and since $B_s = B_0$ without regard to sign

$$j_s = \frac{3B_0 \sin \theta}{2\mu_0} = \frac{3}{2} H_0 \sin \theta \quad . \quad . \quad (11)$$

The vector potential of j_s is circular, concentric with the current. At the surface, where j_s flows, it is given by

$$2\pi p A_s = \pi p^2 B_0$$

or

$$A_s = \frac{1}{2} B_0 p = \frac{1}{2} B_0 R \sin \theta \quad . \quad . \quad (12)$$

The velocity of the effective conduction electrons which constitute the surface current can now be obtained. As in the case of two coupled circuits, v will be the same as if B_s existed by itself; so from eqn. (1)

$$v = \frac{e}{m} A_s = \frac{1}{2} \left(\frac{e}{m} \right) B_0 R \sin \theta \quad . \quad . \quad (13)$$

If the required value of v is greater than that corresponding to the maximum energy which an electron can attain, the current cannot be generated and the field B_0 will penetrate the interior of the sphere.

Let ρ_s be the surface charge density of effective conduction electrons; then

$$\rho_s = \frac{j_s}{v} = \frac{m}{e} \frac{3}{\mu_0 R} \quad . \quad . \quad . \quad (14)$$

The effective conduction charge in the elementary ring of current is therefore

$$\rho_s 2\pi R^2 \sin \theta \delta \theta \quad . \quad . \quad . \quad (15)$$

and dividing by e the number of effective conduction electrons is

$$n = \frac{6\pi m R \sin \theta \delta \theta}{\mu_0 e^2} \quad . \quad . \quad . \quad (16)$$

(3.2) Magnetic Energy of the Surface Current

The kinetic energy of the effective conduction electrons in the elementary ring of current is

$$\frac{1}{2} n m v^2 = \frac{3\pi R^3 B_0^2}{4\mu_0} \sin^3 \theta \delta \theta \quad . \quad . \quad . \quad (17)$$

and integrating from $\theta = 0$ to $\theta = \pi$ gives the kinetic energy of the surface current on the whole sphere as

$$W_m = \pi R^3 (H_0 B_0) \quad . \quad . \quad . \quad (18)$$

This should be equal to the volume integral of $\frac{1}{2} H_s B_s$ for the whole of the field B_s . The magnetic energy of B_s inside the sphere is

$$W_1 = \frac{4}{3} \pi R^3 \left(\frac{H_s B_s}{2} \right) = \frac{2}{3} \pi R^3 (H_0 B_0) \quad . \quad . \quad (19)$$

and the magnetic field of j_s outside the sphere is the same as that of a magnetic dipole of moment $\frac{1}{2} H_s R^3$ at the centre of the sphere. The field at a point r, θ , when $r > R$, has components

$$\left. \begin{aligned} H_{sr} &= H_s \left(\frac{R}{r} \right)^3 \cos \theta \\ H_{s\theta} &= \frac{H_s}{2} \left(\frac{R}{r} \right)^3 \sin \theta \end{aligned} \right\} \quad . \quad . \quad . \quad (20)$$

Consider an elementary ring of space, concentric with the axis of the current, whose position is given by r, θ and whose cross-sectional area is $r \delta \theta \delta r$. The volume of this ring is $2\pi r^2 \sin \theta \delta \theta \delta r$, and the magnetic energy in it has density

$$\frac{1}{2} \mu_0 H_s^2 = \frac{\mu_0}{2} (H_{sr}^2 + H_{s\theta}^2) = \frac{\mu_0 H_s^2 R^6}{2r^6} \left(\cos^2 \theta + \frac{1}{4} \sin^2 \theta \right) \quad (21)$$

The field energy in the ring is therefore

$$\delta W_2 = \frac{\mu_0 \pi H_s^2 R^6}{r^4} \left(\cos^2 \theta \sin \theta + \frac{1}{4} \sin^3 \theta \right) \delta \theta \delta r \quad . \quad (22)$$

and the total external field energy of j_s is

$$W_2 = \int_R^\infty \int_0^\pi dW_2 = \frac{1}{3} \pi R^3 H_s B_s = \frac{1}{3} \pi R^3 (H_0 B_0) \quad . \quad (23)$$

The whole of the magnetic field energy of j_s is therefore

$$W = W_1 + W_2 = \pi R^3 (H_0 B_0) \quad . \quad . \quad (24)$$

which is the same as the kinetic energy of the effective conduction electrons, W_m , as given by eqn. (18).

(3.3) Radial Equilibrium of the Conduction Electrons

The steady surface current j_s consists of electrons moving in free circular paths of radius p . Since the conductor has no resistivity and the macroscopic internal electric field is zero, the current can persist only if the electrons move in the required paths under the action of magnetic forces.

The radial force on a conduction electron due to its motion through the external component B_0 of the magnetic field is

$$f_0 = ev B_0 \quad . \quad . \quad . \quad (25)$$

but from eqn. (13)

$$e B_0 = 2mv/p$$

so that

$$f_0 = \frac{2mv^2}{p} \quad . \quad . \quad . \quad (26)$$

radially inwards.

To find the force arising from the current j_s itself, i.e. from the component of field B_s , let the mass of the surface charge per unit area, ρ_s , be M . Then the contribution of the elementary ring of current to the magnetic energy of j_s is equal to its kinetic energy, i.e.

$$w_m = \frac{1}{2} (2\pi p R \delta \theta) M v^2 = \pi M R v^2 p \delta \theta \quad . \quad (27)$$

The total radial force on the ring of current due to its own field

$$F_p = \frac{\partial w_m}{\partial p (j_s \text{ constant})} \quad . \quad . \quad . \quad (28)$$

and since j_s is to be kept constant in differentiating, we may put $Mv = C$, a constant, so that

$$F_p = \frac{\partial}{\partial p} (\pi C R v p \delta \theta) = 2\pi M R v^2 \delta \theta \quad . \quad . \quad (29)$$

since, from eqn. (13), $\frac{\partial}{\partial p} (vp) = 2v$. Dividing F_p by the area of the surface of the ring, $2\pi p R \delta \theta$, we obtain the radial force on a charge of mass M ,

$$f = \frac{Mv^2}{p}$$

so that the radial force on a conduction electron, due to the field component B_s , is

$$f_s = \frac{mv^2}{p} \quad . \quad . \quad . \quad (30)$$

outwards.

The total radial magnetic force on the electron is therefore

$$f_0 - f_s = \frac{mv^2}{p} \quad . \quad . \quad . \quad (31)$$

inwards, which balances the centrifugal force mv^2/p radially outwards. The effective conduction electrons constituting the surface current are therefore in radial equilibrium.

Furthermore, this equilibrium is stable. Inside the sphere, up to the inner edge of the surface layer of current, $B_s = -B_0$ and the mean value of the axial component of B_s inside the thin current layer itself, acting on the current is $-\frac{1}{2} B_0$ since the force $f_s = -\frac{1}{2} f_0$. So if an electron deviates slightly inwards from its circular path, it moves into a region of greater B_s and the outward force f_s increases, correcting the deviation, while if it deviates slightly outwards, f_s decreases.

(4) THE MEISSNER EFFECT

Consider a solid spherical non-magnetic conductor at a normal temperature situated in a uniform constant magnetic field. According to conventional electromagnetic theory, no action takes place and the magnetic field, regarded macroscopically, is

not disturbed by the conductor; this is consistent with experience. Suppose now that the sphere is cooled until it becomes superconducting and its resistivity disappears. This reduction of the resistivity to zero is the only change which can be recognized by conventional macroscopic theory, and again the magnetic field should not be disturbed, penetrating the conductor as before. If, however, the field is then altered, by eqn. (7) the field inside the conductor should remain constant. This conclusion means, in fact, that such a superconductor should be able to maintain a steady supercurrent of any value even when there is no applied magnetic field, for the latter could be completely removed leaving the supercurrent to maintain an internal field of the original value.

For some 20 years after the discovery of superconductivity by Kamerlingh Onnes³ in 1911, it was assumed that superconductors behave in this manner. Experiments by Meissner and Ochsenfeld⁴ in 1933, however, showed that this is not so, at least for pure superconductors. They found that, when a specimen becomes superconducting in a constant magnetic field, the normal component of B at the surface of the conductor vanishes, showing that the internal field is removed.

For a pure superconductor in an external magnetic field it therefore appears that the internal magnetic field is always zero, no matter how the essential state of affairs is reached; i.e. it does not matter whether a magnetic field is applied to a conductor which is already superconducting, or whether a conductor is made superconducting in an existing magnetic field. In either case the cancellation of the applied field within the conductor must be due to the generation of a surface 'supercurrent' whose value is a function of the applied field. However, if the applied field is too strong, the generation of the supercurrent fails to take place and the field penetrates the conductor. This critical field is found to be a function of temperature for a given superconductor, and may possibly depend on other variables as well.

The discovery of the Meissner effect led to the view that superconductors lie outside the sphere of classical electromagnetic theory. If, however, a conductor with zero resistivity is considered to be a body in which at least a proportion of the conduction electrons move about, with a range of velocities, in free paths without colliding with the lattice of atoms, then it is not difficult to obtain a rough picture of how the effect arises. When a sphere is at a normal temperature and in a magnetic field, all the conduction electrons must suffer collisions after moving short distances in the inter-atomic space, so that although the forces which they experience owing to their motion in the magnetic field may affect paths between collisions, they do not prevent the electrons from having random motions. The electrons cannot therefore be sorted out into a steady macroscopic current. This would require a steady electromotive force to overcome the resistance, and since the applied magnetic field is constant, there is no source of such an e.m.f.

If the conductor is cooled and becomes superconducting at a few degrees absolute, the resistance to steady currents disappears and an induced surface current can, under certain conditions, flow indefinitely. In the transition from the normal to the superconducting state we may suppose that a proportion of the conduction electrons become completely immune from collisions with the lattice and hence can move in free paths through the atomic structure. Owing to the applied magnetic field, these paths will be curved, and some of the electrons will settle into free stationary orbits. They will then form steady currents whose magnetomotive force is easily seen to oppose the applied field. For a perfect Meissner effect the cancellation of the internal field must be complete and there can be no internal steady current, since $\mathbf{J} = \text{curl } \mathbf{H} = 0$. Thus some of the free electrons are sorted out into an orderly surface current, leaving a field-free

interior in which other 'superconducting' electrons move in undisturbed random directions.

If the Meissner effect takes place, the present theory by eqn. (13) specifies the velocity of the effective conduction electrons in the surface current, but if conditions are such that this velocity cannot be attained, the Meissner effect will not occur. It would be going beyond the limits of a macroscopic theory, however, to attempt to specify any upper limit for the electron velocity, since there may be complex processes in the transition whereby the energy of an electron may be increased. Moreover, in evaluating v from eqn. (13) it should be noted that m , the effective mass per electron, is not known to any degree of certainty. All one can say is that it probably is equal to or greater than the accepted rest-mass of a free electron.

The Meissner effect, therefore, evidently arises from the behaviour of electrons in the superconducting state, and thus a full physical theory should be sought in quantum electrodynamics rather than in a macroscopic theory. It is, however, easy to show that, given a pure superconductor and not too strong a magnetic field, the absence of the effect would be inconsistent with the present theory. Let us assume that the effect does not take place: then no supercurrent is generated and the internal field remains constant at the value B_0 . Next, let the applied field be removed, so that, since the field inside the sphere cannot change, a supercurrent must be induced which causes an internal field $B_s = B_0$. The values of ρ_s and v will be the same as in the previous analysis, except that in Fig. 2 the current will be in the opposite direction. The electrons which compose it, however, can no longer be in radial equilibrium, for the stabilizing inward force, f_0 , due to the external field is absent. The resultant radial force on a conduction electron is no longer zero but equal to $2mv^2/\rho$, outwards, and it is evident that a steady current could not persist. The only possible stable physical state when B_0 is removed is $B = 0$, and since the internal field cannot have changed during the removal of B_0 , it follows that B must originally have been zero inside the sphere.

(5) PERFECT CONDUCTORS AND SUPERCONDUCTORS

The distinction which we make between a perfect conductor and a superconductor is different from that usually understood. The distinction adopted by workers in superconductivity is that a perfect conductor is one in which the resistivity is zero and in which a Meissner effect should not be expected to take place, i.e. it is based on conventional electromagnetic theory which, following Maxwell, ignores electron inertia. We regard a perfect conductor, however, as a superconductor in which all conduction electrons would be available to form a supercurrent. Such a material apparently does not exist, for in practical superconductors there are conduction electrons which, under the action of high-frequency electromotive forces, form normal currents which generate heat. In the case of a steady supercurrent $E = 0$, and such normal currents cannot flow, since they are 'short-circuited'.

(6) LONDON EQUATIONS OF SUPERCONDUCTIVITY

It is of interest to attempt a more detailed study of the surface current on the sphere. In the above analysis it appears, in a macroscopic sense, as a current sheet of no finite thickness; but if we think in terms of electrons moving through the atomic structure, it is evident that the sheet must have a finite thickness, however small. Within the sheet there must then be a magnetic field given by $\text{curl } \mathbf{H} = \mathbf{J}$, where \mathbf{J} is the ordinary current density. In the interior of the conductor the field of the supercurrent cancels B_0 completely, so that within the current layer itself this cancellation will be incomplete.

The vector potential, A_0 , of the external uniform field can be chosen so that it is concentric with the axis of the sphere. Its scalar value is then given by

$$A_0 = \frac{1}{2} B_0 \rho \quad (32)$$

in a direction opposed to A_s . At the inner edge of the current layer complete cancellation of B_0 by B_s commences, so $A_s = \frac{1}{2} B_0 \rho$ as given by eqn. (12) and the resultant vector potential is

$$A = A_0 - A_s = 0 \quad (33)$$

We may evidently write these relations vectorially

$$\left. \begin{aligned} A_s &= \frac{m}{e} v_p \\ A_0 &= -\frac{m}{e} v_p \\ A &= A_0 + A_s = 0 \end{aligned} \right\} \quad (34)$$

where $v_p = \frac{1}{2}(e/m)B_0\rho$. As we enter the current layer from the interior of the sphere, v will become progressively less than the value, v_p , necessary for complete cancellation; so suppose that

$$v = v_p - v_d \quad (35)$$

where v_d is some function of the radial distance of the point from the inner edge of the layer where $v_d = 0$. It follows that within the layer

$$A_s = \frac{m}{e} v = -A_0 - \frac{m}{e} v_d \quad (36)$$

so that the resultant vector potential is

$$A = A_0 + A_s = -\frac{m}{e} v_d \quad (37)$$

The magnetic field within the layer is

$$B = \text{curl } A = -\frac{m}{e} \text{curl } v_d \quad (38)$$

In order to obtain an equation describing the distribution of current through the layer, it is necessary to relate v_d to the current density, J . It is doubtful whether this is rigorously possible in a macroscopic theory, but since both quantities are zero at the inner edge of the layer a possible assumption is that v_d and J are directly proportional.

$$\text{If, therefore,} \quad J = \frac{m}{e\Lambda} v_d \quad (39)$$

where Λ is some constant for the material, it follows that

$$B = -\Lambda \text{curl } J \quad (40)$$

$$\text{and thus} \quad \mu_0 J = \text{curl } B = -\Lambda \text{curl curl } J$$

$$\text{or} \quad \text{curl curl } J + \frac{\mu_0}{\Lambda} J = 0 \quad (41)$$

The case considered—that of a solid sphere—is an example of a ‘singly connected’ superconductor. If a supercurrent is induced in a closed ring, eqn. (37) is not valid without some modification, since v_d is zero in the interior of the metal and the line integral of A around a path inside the metal and passing once round the ring must give the linked magnetic flux. Eqn. (37) may be extended to cover this case, that of a ‘doubly connected’ superconductor, by taking eqn. (38):

$$\text{curl } A = -\frac{m}{e} \text{curl } v_d$$

$$\text{whence} \quad A = -\frac{m}{e} v_d + \text{grad } \psi \quad (42)$$

where ψ is a multi-valued potential function similar to the multi-valued magnetic potential of a current circuit. The line

integral of A taken around an interior path where $v_d = 0$ is then $\oint \text{grad } \psi = \Phi$, the linked magnetic flux. Eqn. (36) must also be put in the general form

$$A_s = \frac{m}{e} v = -A_0 - \frac{m}{e} v_d + \text{grad } \psi \quad (43)$$

Eqn. (41) is a basic equation of the London theory of superconductivity.^{5,6} The constant Λ then has the value

$$\Lambda = \frac{m}{ne^2} \quad (44)$$

where n is the total number of available superconducting electrons per unit volume. But v_d in eqns. (39) and (42) appears, not as a very small fraction of the true electron velocity v , but as the true electron velocity itself, denoted in the London theory by v_s . Eqn. (39) then becomes $J = nev_s$ and eqn. (42) appears in the form

$$p_s = mv_s + eA = \text{grad } \chi \quad (45)$$

where p_s is the ‘momentum’ of the supercurrent and χ is the ‘superpotential’. Therefore

$$\frac{\partial p_s}{\partial t} = m \frac{\partial v_s}{\partial t} + e \frac{\partial A}{\partial t} \quad (46)$$

It is supposed that the acceleration of the electrons is produced by the resultant electric field, as in the ‘acceleration theory’ of Becker, Heller and Sauter,⁷ so that

$$m \frac{\partial v_s}{\partial t} = eE \quad \text{and} \quad E = \frac{m}{e} \frac{\partial v_s}{\partial t} \quad (47)$$

Hence, from eqn. (46)

$$E = -\frac{\partial A}{\partial t} + \frac{1}{e} \frac{\partial p_s}{\partial t} \quad (48)$$

and since $E = -\frac{\partial A}{\partial t} - \text{grad } \phi$ it follows that

$$\frac{\partial p_s}{\partial t} = -e \text{grad } \phi \quad (49)$$

where $-\text{grad } \phi$ is the electrostatic component of the electric field. The momentum p_s also satisfies the relations

$$\text{curl } p_s = 0, \quad \text{div } p_s = 0 \quad (50)$$

and is zero for a singly-connected superconductor.

The regular solutions of eqn. (41) are such that the current decreases very rapidly as the distance from the surface, towards the interior, increases. Hence this equation is consistent with the Meissner effect and at a depth greater than $\sqrt{(\Lambda/\mu_0)}$ the field and current are practically zero. This penetration depth is of the order of 10^{-6} to 10^{-5} cm.

The fallacy in this ‘acceleration’ theory, on which the London equations are founded, is that the electromagnetic basis of the concept of electronic mass is completely ignored. In the case of a single electron in an electric field E arising from external sources, eqn. (47) is valid and is equivalent to

$$e(E + E_s) = 0 \quad (51)$$

which states that the electron, regarded as a charge of no inertia apart from that of its electromagnetic field, moves in such a way that the total electromagnetic force acting on it is zero. The component E_s is the field, averaged over the charged particle, induced by its own motion and, for the quasi-stationary state, is given by

$$eE_s = -m \frac{\partial v_s}{\partial t} \quad (52)$$

In the London theory, however, v_s is not the random velocity of a single electron in an ‘electron gas’, but the velocity at a point in the aggregate stream of conduction electrons in their

ordered motion, as is obvious from the relation $\mathbf{J} = nev_s$. The microscopic concept of discrete charged particles can, in fact, be replaced in a macroscopic theory by that of the steady flow of an electric fluid in which \mathbf{v}_s is the velocity field. The ratio m/e is thus the ratio of electromagnetic mass to charge of the aggregate and so, unless m is to be taken as being non-electromagnetic, eqn. (52) applies to the aggregate stream and \mathbf{E}_s is the macroscopic average self-induced electric field of the changing current. Thus if we adhere to this accepted meaning of electronic mass, eqn. (47) actually states

$$-\mathbf{E}_s = \mathbf{E} = \mathbf{E}_0 + \mathbf{E}_s$$

$$\text{or} \quad -\mathbf{E}_s = \frac{1}{2}\mathbf{E}_0 \quad (53)$$

where \mathbf{E}_0 is the component induced by the external source.

Now if the density of the magnetic flux linking the current has components \mathbf{B}_0 from the external source and \mathbf{B}_s from the supercurrent, neglecting the very small flux within the thin surface current layer it follows that

$$\mathbf{B}_0 + \mathbf{B}_s = \text{a constant}$$

$$\text{so that} \quad \frac{d}{dt} \iint \mathbf{B}_s dS = - \frac{d}{dt} \iint \mathbf{B}_0 dS$$

$$\text{and therefore} \quad - \oint \mathbf{E}_s dl = \oint \mathbf{E}_0 dl \quad (54)$$

which is incompatible with eqn. (53).

Nevertheless, the London equations describing the supercurrent itself have proved of considerable value as a phenomenological, or—in the language of engineering science—an empirical theory, in which the aim is merely to find equations, regardless of fundamental principles, which describe what actually happens. It now appears, however, that more refined measurements of the penetration depth in different superconducting metals and with varying degrees of impurity, such as in the experiments of Pippard^{8,9} and Faber,⁹ indicate that the theory is not altogether satisfactory, even from this limited viewpoint.

In terms of the theory presented in this paper the problem of the depth of penetration of a supercurrent is microscopic rather than macroscopic, i.e. it is a problem for quantum electrodynamics to see where the superconducting electrons with the necessary energy and freedom may be found.

(7) INERTIAL SUPERCURRENT IN A ROTATING SPHERE

From eqn. (14), when a supercurrent flows in the surface layer of a sphere it is due to the motion of a characteristic charge whose surface density is uniform over the sphere, being a function of m/e and the radius of the sphere. From eqn. (13) it is also evident that this surface charge rotates rigidly about the axis. Thus, disregarding the composition of this surface charge in terms of electrons, we may think of a superconducting sphere as possessing an extremely thin shell of negative charge, of uniform density, having inertia but completely free to rotate about any diameter of the sphere without resistance to motion or electromagnetic reactions. If the sphere is in the superconducting state with no applied field, this shell can therefore be set in rotation only by an electric field induced by external sources.

Suppose the sphere to be stationary with no applied field, and then to be set into uniform rotation about a diameter. The spherical shell of negative charge will remain stationary, owing to its inertia, and the rotation of the equal and opposite positive charge in the surface layer will constitute a surface current, causing a magnetic field.

Let ω be the angular velocity of the sphere. Then the surface current density will be

$$j_s = \rho_s \omega p = \frac{m}{e} \frac{3\omega \sin \theta}{\mu_0} \quad (55)$$

from eqn. (14). The relation between j_s and the internal magnetic field is given by eqn. (11), so that

$$\frac{3}{2} \frac{B}{\mu_0} \sin \theta = \frac{m}{e} \frac{3\omega}{\mu_0} \sin \theta$$

$$\text{and} \quad B = 2\omega \frac{m}{e} \quad (56)$$

inside the sphere. It is of interest to note that this is identical with the result obtained, by somewhat lengthy analysis, from the London theory.¹⁰

(8) ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. D. Shoenberg and Dr. A. B. Pippard of the Royal Society Mond Laboratory, Cambridge, for helpful comments and criticism, and to his colleague Dr. D. Midgeley, who contributed to Section 3.2.

(9) REFERENCES

- (1) CULLWICK, E. G.: 'Electromagnetic Momentum and Electron Inertia in a Current Circuit', *Proceedings I.E.E.*, Monograph No. 150, September, 1955 (103 C, p. 159).
- (2) MAXWELL, J. C.: 'A Treatise on Electricity and Magnetism' (University Press, Oxford, Third Edition, 1892), Volume II, Part IV, §655, p. 290.
- (3) KAMERLINGH ONNES, H.: 'Further Experiments with Liquid Helium. D. On the Change of the Electrical Resistance of Pure Metals at very low Temperature, etc. V. The Disappearance of the Resistance of Mercury', *Communications from the Physical Laboratory of the University of Leiden*, 1911, 122b, p. 13. 'VI. On the Sudden Change in the Rate at which the Resistance of Mercury Disappears', *ibid.*, 1911, 124c, p. 21. 'Untersuchungen über die Eigenschaften der Körper bei niedrigen Temperaturen, welche Untersuchungen unter anderem auch zur Herstellung von flüssigem Helium geführt haben', *ibid.*, 1913, Supplement 35, p. 1.
- (4) MEISSNER, W., and OCHSENFELD, R.: 'Ein neuer Effekt bei Eintritt der Supraleitfähigkeit', *Naturwissenschaften*, 1933, 21, p. 787.
- (5) LONDON, F., and LONDON, H.: 'Electromagnetic Equations of the Superconductor', *Proceedings of the Royal Society, A*, 1935, 149, p. 71.
- (6) LONDON, F.: 'Superfluids. Volume I: Macroscopic Theory of Superconductivity' (Wiley, New York, 1950).
- (7) BECKER, R., HELLER, G., and SAUTER, F.: 'Über die Stromverteilung in einer Supraleitenden Kugel', *Zeitschrift für Physik*, 1933, 85, p. 772.
- (8) PIPPARD, A. B.: 'An Experimental and Theoretical Study of the Relation between Magnetic Field and Current in a Superconductor', *Proceedings of the Royal Society, A*, 1953, 216, p. 547.
- (9) FABER, T. E., and PIPPARD, A. B.: 'The Penetration Depth and High-Frequency Resistance of Superconducting Aluminium', *Proceedings of the Royal Society, A*, 1955, 231, p. 336.
- (10) LONDON, F.: 'Superfluids, Volume I' (Wiley, New York, 1950), pp. 78–82.

THE POLARIZATION OF VERY LONG RADIO WAVES REFLECTED FROM THE IONOSPHERE AT OBLIQUE INCIDENCE

By W. C. BAIN, M.A., B.Sc., Ph.D., and C. B.-I. GLASS, B.Sc., Ph.D.

(The paper was first received 17th February, and in revised form 16th April, 1956. It was published as an INSTITUTION MONOGRAPH June, 1956.)

SUMMARY

A study of some additional evidence has been made to determine the polarization of radio waves reflected from the ionosphere at 16 kc/s over a transmission path 540 km long. The ratio between the amplitudes of the vertical and horizontal electric fields is found to be appreciably greater than unity, the mean value of the determinations made here being 5.

(1) INTRODUCTION

In a previous paper Bain, Bracewell, Straker and Westcott* gave figures for the state of polarization of the radio waves received at Aberdeen from the transmitter GBR, which radiates at 16 kc/s from a site near Rugby. The method of measuring this quantity was to record the phase and amplitude of the received signals on loop aerials in and at right angles to the plane of propagation. When the recorded phase and amplitude are plotted as polar co-ordinates, the resulting points in either case lie on a characteristic curve (the loop-induction locus) produced by the variation of the signal during the day. The respective loci will have similar shapes if the polarization remains unchanged throughout the day.

Now suppose that the ground wave at the receiving aerial has a magnetic vector H_0 . Let the normal component of the downcoming wave have a magnetic vector H_1 (at right angles to the plane of propagation) and let the magnetic vector of the abnormal component be H_2 (in the plane of propagation). With the assumption, for the moment, of a single downcoming wave, the resultant horizontal magnetic field components at the receiving site are given by

$$H_A = 2H_2 \cos i \quad (1)$$

$$H_N = H_0 + 2H_1 \quad (2)$$

where H_A is in the direction of propagation, H_N is at right angles to this direction and i is the angle of incidence of the downcoming wave at the receiving aerial.

The shapes of the loci for H_N and H_A given in the previous paper were roughly similar, and it was deduced that $|H_2|/|H_1| = 1$, and that H_2 leads H_1 in phase by 180° .

A re-examination of the original evidence and a study of some additional data have now been carried out, and it has been found that this estimate of the polarization is incorrect. Although both loci show roughly the same total change from night to day conditions, the change takes place more rapidly for H_A than for H_N ; hence the polarization of the downcoming wave is presumably changing, which is contrary to the assumption on which the previous calculation was based. It has also been realized that a very useful opportunity of determining at least the ratio $|H_2|/|H_1|$ is afforded by the fact that during the period

1st April–15th October, approximately, H_N falls to a very low value at the times of rapid change of signal in both the morning and the evening. This means that the measured values of H_A at these times are not upset by incorrect loop setting; for even if the loop is not adjusted to suppress H_N completely, the undesired component of H_N in the nominal value of H_A is of negligible amplitude.

(2) RESULTS

A number of records which can be treated in this way have been obtained at Aberdeen during the summer months. Six of these records were taken by the process of simultaneous measurement of phase and amplitude with loops set parallel to and perpendicular to the plane of propagation, and another two by measurements of radio bearing with manual rotation of a loop aerial. To find the polarization from these observations the values of $2H_1$ and $H_A (=2H_2 \cos i)$ have to be found at the time of minimum H_N . Then, if a value for $\cos i$ is assumed, the ratio $|H_1|/|H_2|$ can be calculated.

The value of H_A can be read off directly as the amplitude of the received signal on the loop set at right angles to the plane of propagation at the time when H_N is a minimum. To find $2H_1$ at this time from the phase and amplitude records, however, requires a knowledge of the ground-wave vector H_0 . This can be obtained by a method described by Bain *et al.*, and $2H_1$ can then be found from the relation

$$2H_1 = H_N - H_0$$

A somewhat different procedure is necessary in dealing with the radio-bearing results, which were not accompanied by phase measurements. Here the value of $|H_0|$ is obtained from

$$|H_0| = 2 \cdot 02 |H_N|$$

where $|H_N|$ is now the value at summer noon. This relationship can be deduced from the figures given by Bain *et al.* for the amplitude and phase relations between the ground wave, the sky wave and the resultant signal at this time. The value of $|2H_1|$ at minimum H_N can then be derived from

$$|2H_1| = |H_N - H_0|$$

an equation which holds at this time as H_0 and $2H_1$ are then in phase opposition.

To obtain H_2 from eqn. (1), $\cos i$ has been taken to be 0.29, corresponding to an angle of incidence of 73° for the downcoming wave and to an apparent height of reflection of about 75 km. The results of the polarization calculations are summarized in Table 1.

It is clear that the ratio of $|H_1|/|H_2|$ is much greater than the result quoted in the previous paper, the mean of the tabulated values being 5. The variability of the results is no doubt largely due to inaccuracies in the measurement of H_A , which was always of very small amplitude at the times in question.

* BAIN, W. C., BRACEWELL, R. N., STRAKER, T. W., and WESTCOTT, C. H.: 'The Ionospheric Propagation of Radio Waves of Frequency 16kc/s over Distances of about 540 km', *Proceedings I.E.E.*, Monograph No. 37, May, 1952 (99, Part IV, p. 250).

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Bain is at the Radio Research Station, Department of Scientific and Industrial Research, and Dr. Glass is in the Royal Naval Scientific Service.

Table 1
MAGNETIC-FIELD COMPONENTS OF GBR AT ABERDEEN AT THE
TIME OF MINIMUM H_N

(Note.—The units of field are arbitrary and differ from day to day)

Date	Time	$ H_N $	$ 2H_1 $	$ H_A $	$ H_1 / H_2 $
	U.T.				
6th September, 1949	0509	1.5	5.3	0.5	3.1
6th September, 1949	1902	0.5	3.5	0.4	2.6
7th September, 1949	0512	1.4	6.5	0.55	3.5
7th September, 1949	1901	0.2	5.5	0.4	4.0
3rd May, 1950 ..	0413	0.0	6.5	0.5	3.7
1st September, 1950	0436	2.1	11.7	0.4	8.5
4th June, 1951 ..	2045	0.02	0.53	0.015	10.1
27th June, 1951 ..	2053	0.06	0.65	0.03	6.4

In reaching the above conclusion regarding polarization, the presence of only a single downcoming ray has been assumed. The figures might therefore be in error if the low value of H_A obtained could be ascribed to the fortuitous cancellation of two or more vectors due to waves of considerable amplitude reflected more than once from the ionosphere. However, the phases of all components are altering rapidly near the time of minimum H_N , and if the low value of H_A were due to such cancellation, a

large abnormal component would appear shortly before and after this time. Since this is contrary to observation, the previous conclusion remains valid.

The polarization at other times of the summer day is more difficult to determine. It is found that the radio bearing remains sharp and constant during the time when the zenith distance of the sun is less than 92° , and this shows that there is then no great increase in the H_2 component. When the zenith distance exceeds 98° there is a comparatively large horizontally polarized component; but at such times a considerable part of the signal is due to multiple reflections, and the proportion of once-reflected wave in this component has not been determined. Similar remarks apply to the signal at all times of a winter day. Other comments on this question have been given by Bracewell and Bain.*

(3) CONCLUSIONS AND ACKNOWLEDGMENTS

In the downcoming wave at Aberdeen from the station GBR the ratio between its vertically and horizontally polarized components is considerably greater than unity, the mean value being 5.

The observational work described here was carried out as part of the research programme of the Natural Philosophy Department, University of Aberdeen.

* BRACEWELL, R. N., and BAIN, W. C.: 'An Explanation of Radio Propagation at 16 kc/s in Terms of Two Layers below E-Layer', *Journal of Atmospheric and Terrestrial Physics* 1952, 2, p. 216.

A STUDY OF THE FIELD DISTRIBUTION AT AN AXIAL FOCUS OF A SQUARE MICROWAVE LENS

By P. A. MATTHEWS, B.Sc., Student, and Professor A. L. CULLEN, Ph.D., Associate Member.

(The paper was first received 6th February, and in revised form 5th April, 1956. It was published as an INSTITUTION MONOGRAPH in July, 1956.)

SUMMARY

The field distribution near the focus of a microwave lens is studied both theoretically and experimentally.

The relation between vector-field diffraction theory and scalar-field diffraction theory is discussed for this particular problem, and it is shown that in the case in question the principal component of the electric field vector can be evaluated with sufficient accuracy from the scalar theory.

Measurements of the transverse component of the electric field have been made by a perturbation method using a spinning dipole. The experimental results are in good agreement with the theory.

The often-discussed phase-shift of 180° in the passage of a wave through the focal plane is considered in some detail and the related change in wavelength near the focus is verified experimentally.

LIST OF PRINCIPAL SYMBOLS

- a = A parameter specifying parabolic taper.
 $A(S)$ = Aperture distribution function.
 $C(x)$ = Real part of Fresnel integral.
 E = Electric field vector.
 H = Magnetic field vector.
 k = Phase constant = $2\pi/\lambda$.
 n = Distance measured in the direction of the normal to a surface.
 \mathbf{n} = Unit vector normal to a surface.
 r, R = Distances.
 $S(x)$ = Imaginary part of Fresnel integral.
 x, y, z = Cartesian co-ordinates.
 α, β = Angles related to ray directions.
 ϵ = Permittivity of medium.
 ϕ = Scalar wave function.
 $\psi = e^{-jkr}/r$.
 λ = Wavelength.
 μ = Permeability of the medium.

(1) INTRODUCTION

The determination of the field distribution near the focus of an incoming spherical wave has been of interest for many years and has received extensive theoretical study.

In 1834 Airy¹ derived an expression for the intensity in the focal plane produced by a converging spherical wave limited in extent by a circular aperture. At that time the electromagnetic theory of light had not been formulated and Airy used the scalar wave theory in his calculations.

Lommel,² in 1886, gave a more general solution which made it possible to calculate the field at points near the focus, but not necessarily in the focal plane.

The occurrence of a phase shift of 180° on the passage of the wave through the focal plane was pointed out by Gouy³ in 1890, and he showed that this was associated with an increase

in phase velocity near the focus. Several other papers^{4,5,6} on the same topic appeared at about that time.

The first attempt to treat the problem as one of electromagnetic wave theory was made in 1909 by Debye,⁷ who used the Hertzian vector method. He showed that the vector field could be evaluated from one rectangular component of the Hertzian vector and so reduced the problem to a scalar diffraction calculation. He also showed that the field along the axis could be evaluated in terms of Fresnel's integrals.

All these investigators treated the problem from the standpoint of the optics of visible light, at a time when the intensity distribution was of main interest. In microwave optics, intensity, phase and polarization are all of interest and can all be measured because coherent monochromatic sources are available. The longer wavelength compared with that of visible light makes it comparatively easy to measure the fine structure of the diffraction pattern. It is therefore possible to study experimentally in considerable detail the field distribution near a focal point in an electromagnetic field, and to compare the results with the theory. This has been the objective in the work to be described.

(2) THEORY OF FOCUSING

The theory of the field distribution will first be considered. It will be necessary to allow for a tapered aperture distribution, and consideration must be given to the vector character of the field. It has been thought preferable to avoid introducing the Hertzian vector, and it has been verified directly in Section 2.3 that the scalar theory can be used to calculate the principal component of the electric field vector with sufficient accuracy for the purpose in view.

(2.1) Fundamental Diffraction Theory

The diffraction theory of physical optics has developed from Huygens's principle. Fresnel first formulated the principle mathematically, but it was left to Helmholtz to derive a rigorous diffraction formula taking the scalar wave equation as a starting-point. Helmholtz considered monochromatic waves only; the theory for arbitrary time-dependence was developed later by Kirchhoff. A completely analogous procedure for the vector wave equation derived from Maxwell's equations of the electromagnetic field has been worked out by Stratton and Chu.⁸

It is unnecessary to consider arbitrary time-dependence here, but the vector nature of the field must be borne in mind. However, the scalar wave theory will be considered first and its application to a vector-field problem deferred for treatment later.

Let ϕ be a scalar wave function which at any point in space satisfies the scalar wave equation

$$\nabla^2 \phi + k^2 \phi = 0 \quad (1)$$

Consider a volume V within an isotropic homogeneous medium bounded by a closed surface S (see Fig. 1). If ϕ is continuous and has continuous first derivatives within V and on S , it follows from Green's theorem that the value of ϕ at an interior point x', y', z' can be expressed as an integral involving ϕ and its

Correspondence on Monographs is invited for consideration with a view to publication.

Prof. Cullen is Professor of Electrical Engineering, University of Sheffield. Mr. Matthews is at the Radio Research Station, Department of Scientific and Industrial Research.

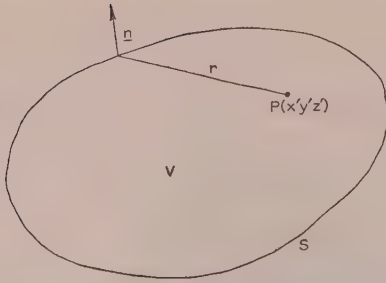


Fig. 1.—Volume V enclosed by a surface S and containing the point of observation P .

normal derivative taken over the surface S . The value of ϕ is given by

$$4\pi\phi = \int_S \left[\frac{\partial\phi}{\partial n} \frac{e^{-jkr}}{r} - \phi \frac{\partial}{\partial n} \left(\frac{e^{-jkr}}{r} \right) \right] dS \quad (2)$$

where r is the distance from a variable point on S to the fixed interior point x', y', z' , and n is the distance measured along the outward normal to the surface S .

This is the formula obtained by Helmholtz for the case of sinusoidal time variation. A time factor $e^{j\omega t}$ is assumed in eqns. (1) and (2).

Now assume that the surface S is an opaque screen separating a source from the observer. The observer is assumed to be inside the volume V , but this does not restrict his position since the surface S may be assumed to be closed by a surface of infinite extent. If an aperture S_1 is made in the screen the field will penetrate into the region occupied by the observer. To determine the intensity and distribution of the field inside V the values of ϕ and of $\partial\phi/\partial n$ must be known over the aperture and over the observer's side of the screen. Usually these values are not known, but in order to obtain an approximate solution it may be assumed that

- (a) On the surface of the screen $\phi = 0$, $\partial\phi/\partial n = 0$.
- (b) Over the surface S_1 , the values of ϕ and $\partial\phi/\partial n$ are those which would exist if no screen were present.

When these approximations are made the diffracted field may be calculated.

This procedure gives results which are in reasonable agreement with experiment, provided that the aperture is large compared with the wavelength. For a small aperture the discontinuity at the edge of the aperture modifies the diffracted field profoundly. As the area of the aperture increases, the relative importance of the contribution from the edge often decreases, and for sufficiently large apertures it may be negligible. Although results based on the assumptions are usually satisfactory, they are strictly inconsistent with assumption (a), for, by a theorem in the theory of functions, if ϕ and $\partial\phi/\partial n$ are zero over any finite part of S they are zero at all points in the space enclosed by S . Moreover, the two assumptions are mutually inconsistent for the same reason. Very illuminating discussions of this difficulty have been given by Sommerfeld⁹ and Schelkunoff.¹⁰

(2.2) Application of Scalar Diffraction Theory to the Calculation of the Field Distribution near a Focal Point

In Section 2.1 it has been shown that the wave function ϕ at any point P (Fig. 1) can be calculated in terms of the values of ϕ and $\partial\phi/\partial n$ on a surface enclosing P .

This result will be applied to the idealized situation indicated in Fig. 2. The volume V containing the point of observation P

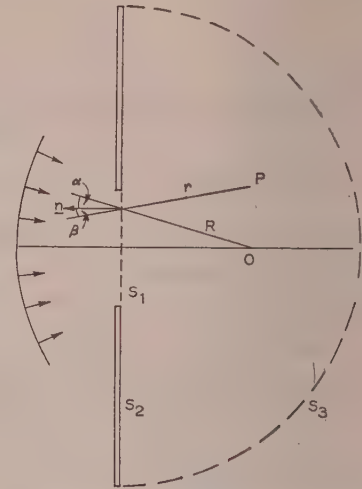


Fig. 2.—Diffracting screen with incident spherical wave.

is bounded by a surface S which is conveniently subdivided in the following way. S_1 is a plane surface in the aperture of the infinite plane screen, and S_2 is the opaque portion of this screen. S_3 is the surface of a hemisphere of infinite radius with its centre in the aperture.

The incident wave is taken to be a converging spherical wave which in the aperture plane has the form

$$\phi = \frac{e^{jkR}}{R} \quad (3)$$

With the notation of Fig. 2 it is easily shown that

$$\frac{\partial\phi}{\partial n} = jk \left(1 - \frac{1}{jkR} \right) \frac{e^{jkR}}{R} \cos \alpha \quad (4)$$

Substituting eqns. (3) and (4) in eqn. (2) and carrying out the remaining differentiations,

$$4\pi\phi = jk \int_S \left[\left(1 - \frac{1}{jkR} \right) \cos \alpha + \left(1 + \frac{1}{jkr} \right) \cos \beta \right] \frac{e^{jk(R-r)}}{Rr} dS \quad (5)$$

If $kR \gg 1$ and $kr \gg 1$, which is certainly true in the experimental work described later,

$$4\pi\phi = jk \int_S (\cos \alpha + \cos \beta) \frac{e^{jk(R-r)}}{Rr} dS \quad (6)$$

This is essentially Fresnel's diffraction formula adapted to the particular problem.

It has been assumed in the derivation of eqn. (6) that the incident wave is isotropic. If this is not so, eqn. (6) can be modified by a factor $A(S)$ in the integrand, which can be chosen to take into account any distribution of amplitude over the aperture. The modified formula is therefore

$$4\pi\phi = jk \int_S A(S) (\cos \alpha + \cos \beta) \frac{e^{jk(R-r)}}{Rr} dS \quad (7)$$

In the experiments described later, a rectangular aperture was used, and it will be convenient to express eqn. (7) in terms of a Cartesian co-ordinate system.

Referring to Fig. 3, consider a rectangular aperture defined by $|\eta| \leq \eta_0$, $|\xi| \leq \xi_0$ in a plane screen. Let Q be a point in

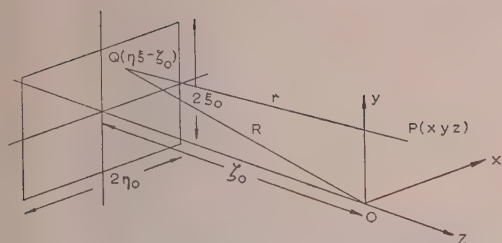


Fig. 3.—Rectangular aperture and co-ordinate system.

the aperture plane, having rectangular co-ordinates $\eta, \xi, -\zeta_0$ in a co-ordinate system whose origin is at the focal point O, and let P be the point of observation, with co-ordinates x, y, z in the same co-ordinate system. The distance PQ is given by

$$PQ^2 = r^2 = (x - \eta)^2 + (y - \xi)^2 + (z + \zeta_0)^2$$

If attention is confined to a region near O, small in comparison with the aperture dimensions, the following approximation holds good:

$$r^2 \simeq R^2 - 2(x\eta + y\xi - z\zeta_0)$$

where

$$R^2 = \eta^2 + \xi^2 + \zeta^2$$

Expanding binomially and again neglecting higher powers of $(x\eta + y\xi - z\zeta_0)/R$, we have

$$PQ = r \simeq R - \frac{1}{R}(x\eta + y\xi - z\zeta_0)$$

$$\text{Thus } R - r \simeq \frac{1}{R}(x\eta + y\xi - z\zeta_0) \quad (8)$$

$$\text{Also } \frac{1}{R} \simeq \frac{1}{\zeta_0} \left(1 - \frac{\eta^2 + \xi^2}{2\zeta_0^2}\right) \quad (9)$$

The maximum value of $z\zeta_0$ is much greater than the maximum value of $x\eta + y\xi$ in the region of interest around the focal point, so that, whilst $1/R$ can be replaced by $1/\zeta_0$ in the evaluation of $(x\eta + y\xi)/R$, the more accurate equation, eqn. (9), must be used in evaluating $z\zeta_0/R$. When this is done,

$$R - r \simeq \frac{x\eta + y\xi}{\zeta_0} - z + \left(\frac{\eta^2 + \xi^2}{2\zeta_0^2}\right)z \quad (10)$$

Eqn. (10) will be used in the phase factor $e^{jk(R-r)}$ in evaluating eqn. (7). However, the amplitude factor in eqn. (7) does not need such careful treatment, and it is permissible to replace Rr by ζ_0^2 and to put both $\cos \alpha$ and $\cos \beta$ equal to unity. When these substitutions are made in eqn. (7), the following result is obtained:

$$\phi = \frac{j\epsilon^{-jkz}}{\lambda\zeta_0^2} \int_{-\zeta_0}^{+\zeta_0} \int_{-\eta_0}^{+\eta_0} A(\eta_1\xi) \exp \left[jk \left(\frac{x\eta + y\xi}{\zeta_0} + \frac{\eta^2 + \xi^2}{2\zeta_0^2} z \right) \right] d\eta d\xi \quad (11)$$

(2.3) Scalar Theory as an Approximation to the Vector Electromagnetic-Field Theory

Before the integration of eqn. (11) is developed its relevance to the microwave experiments will be shown by demonstrating that it gives a good approximation to the principal component E_x of the electric field vector, provided that the distance of the point of observation from the axis in the x -direction is small in comparison with the axial distance ζ_0 between the lens and the focal point, and provided that the aperture distribution is symmetrical.

It may be shown⁸ that if a volume V enclosed by a surface S contains no charge or current within its interior or on its bounding surface, the field at an interior point (x, y, z) is given in terms of the field vectors E and H on its boundary surface by

$$4\pi E = \int_S [j\omega\mu(\mathbf{n} \times \mathbf{H})\psi - (\mathbf{n} \times \mathbf{E}) \times \nabla\psi - (\mathbf{n} \cdot \mathbf{E})\nabla\psi] dS \quad (12)$$

$$4\pi H = \int_S [-j\omega\epsilon(\mathbf{n} \times \mathbf{E})\psi - (\mathbf{n} \times \mathbf{H}) \times \nabla\psi - (\mathbf{n} \cdot \mathbf{H})\nabla\psi] dS \quad (13)$$

where $\psi = e^{-jkr}/r$, and \mathbf{n} is a unit vector in the direction of the outward normal.

In these expressions the values of E and H cannot be specified independently, but must be chosen to satisfy Maxwell's equations. Moreover, they hold only if the vectors E and H are continuous and have continuous first derivatives over S (cf. the similar conditions on ϕ in the scalar case). For diffraction at an aperture bounded by a conducting screen, it is necessary to assume a distribution of line charges and currents along the boundary to satisfy the field equations on passing across the boundary.

One method of finding a boundary distribution which satisfies the field equations has been proposed by Kottler,⁸ who showed that the field at the interior point (x, y, z) is given by

$$4\pi E = \int_S [j\omega\eta(\mathbf{n} \times \mathbf{H})\psi - (\mathbf{n} \times \mathbf{E}) \times \nabla\psi - (\mathbf{n} \cdot \mathbf{E})\nabla\psi] dS + \frac{1}{j\omega\epsilon} \oint_C \nabla\psi \mathbf{H} \cdot d\mathbf{l} \quad (14)$$

$$4\pi H = \int_S [-j\omega\epsilon(\mathbf{n} \times \mathbf{E})\psi - (\mathbf{n} \times \mathbf{H}) \times \nabla\psi - (\mathbf{n} \cdot \mathbf{H})\nabla\psi] dS - \frac{1}{j\omega\mu} \oint_C \nabla\psi \mathbf{E} \cdot d\mathbf{l} \quad (15)$$

where $d\mathbf{l}$ is directed along the boundary contour.

In these expressions the values of E and H are those in the aperture. It is assumed that everywhere else on the surface E and H are both zero.

It may be shown that eqn. (14) is identical with

$$4\pi E = \int_S \left(\psi \frac{\partial E}{\partial n} - E \frac{\partial \psi}{\partial n} \right) dS + \frac{1}{j\omega\epsilon} \oint_C \nabla H \cdot d\mathbf{l} + \oint_C \psi E \times d\mathbf{l} \quad (16)$$

Assume that in the aperture plane

$$\begin{aligned} E &= iE_x \\ H &= jH_y \end{aligned} \quad (17)$$

Then eqn. (16) becomes

$$4\pi E = i \int_S \left(\psi \frac{\partial E_x}{\partial n} - E_x \frac{\partial \psi}{\partial n} \right) dS + \frac{1}{j\omega\epsilon} \oint_C \nabla \psi H_y j \cdot d\mathbf{l} + \oint_C \psi E_x i \times d\mathbf{l} \quad (18)$$

For this to give $E_x(x, y, z)$ identical with the scalar distribution it is necessary to show that the contributions of the two line integrals are negligible.

In the second line integral, $i \times d\mathbf{l}$ is perpendicular to i , so that this integral contributes nothing to the x -component of E .

In the first, we note that

$$\nabla\psi = \frac{\partial}{\partial r} \left(\frac{e^{-jkr}}{r} \right) \frac{\mathbf{r}}{r}$$

where $r = i(x - \eta) + j(y - \xi) + k(z + \xi_0)$

so that $(\nabla\psi)_x = \frac{\partial}{\partial r} \left(\frac{\varepsilon^{-jkr}}{r} \right) \frac{x - \eta}{r}$

or $(\nabla\psi)_x \approx -\frac{\partial}{\partial r} \left(\frac{\varepsilon^{-jkr}}{r} \right) \frac{\eta}{r} \quad (19)$

since $x \ll r$ near the focus, by the previous assumption that $x \ll \xi_0$. In the line integral the only contributions come from the sides of the rectangle parallel to the y -axis.

If attention is confined to symmetrical aperture distributions, H_y is an even function of η . The distance r is also an even function of η if x is negligible in comparison with r , as has been assumed. Thus $\frac{1}{r} \frac{\partial}{\partial r} \left(\frac{\varepsilon^{-jkr}}{r} \right)$ is an even function of η .

It follows from eqn. (19) that $(\nabla\psi)_x H_y$ is an odd function of η . Since η has the value $-\eta_0$ on one of the two sides of the rectangle concerned and $+\eta_0$ on the other, the resultant line integral vanishes. Hence, if terms involving x/r are negligible in comparison with unity, we can write

$$4\pi E_x \int_S \left[\frac{\partial E_x}{\partial n} \left(\frac{\varepsilon^{-jkr}}{r} \right) - E_x \frac{\partial}{\partial n} \left(\frac{\varepsilon^{-jkr}}{r} \right) \right] dS \quad (20)$$

The right-hand side of this equation is identical with the right-hand side of the basic scalar diffraction equation, eqn. (2), and hence eqn. (11), derived from eqn. (2), will give a good approximation to the x -component of the electric field near the focal point if $A(\eta, \xi)$ is interpreted as the x -component of the electric field in the aperture plane.

Further analysis shows that the ratio of axial to transverse electric field strengths is of the order x/r , and so the field may be assumed to be wholly transverse to an accuracy consistent with the other approximations made and the accuracy of measurement.

(2.4) Distribution near the Focus for Two Special Aperture Distributions

Two special cases are now considered: first, a uniform aperture distribution, and second, a parabolic variation of amplitude over the aperture.

(2.4.1) Uniform Aperture Distribution.

For this case let $A(\eta, \xi) = 1$ in eqn. (11); the integrations with respect to η and ξ are then separable; thus

$$\phi = \frac{j \exp(-jkz)}{\lambda \xi_0^2} \int_{-\eta_0}^{+\eta_0} \exp \left[jk \left(\frac{x\eta}{\xi_0} + z \frac{\eta^2}{2\xi_0^2} \right) \right] d\eta \times \int_{-\xi_0}^{+\xi_0} \exp \left[jk \left(\frac{y\xi}{\xi_0} + z \frac{\xi^2}{2\xi_0^2} \right) \right] d\xi \quad (21)$$

This integral can be evaluated by the method devised by Lommel.² The result is

$$\phi(p, p', z) = \frac{jk\eta_0\xi_0}{\xi_0^2} \exp \left[-j \left(kz + \frac{q + q'}{2} \right) \right] \times [U_{1/2}(q, p) + jU_{3/2}(q, p)] \frac{1}{\sqrt{q}} \times [U_{1/2}(q', p') + jU_{3/2}(q', p')] \frac{1}{\sqrt{q'}} \quad (22)$$

where $U_n(q, p) = \sum_{m=0}^{\infty} (-1)^m \left(\frac{q}{p} \right)^{n+2m} J_{n+2m}(p)$

$$\left. \begin{aligned} p &= \frac{kx\eta_0}{\xi_0}, & q &= \frac{kz\eta_0^2}{\xi_0^2} \\ p' &= \frac{ky\xi_0}{\xi_0}, & q' &= \frac{kz\xi_0^2}{\xi_0^2} \end{aligned} \right\} \quad (23)$$

In Fig. 4 contours of equal amplitude, derived from eqn. (22), have been plotted. Experimental results have been obtained

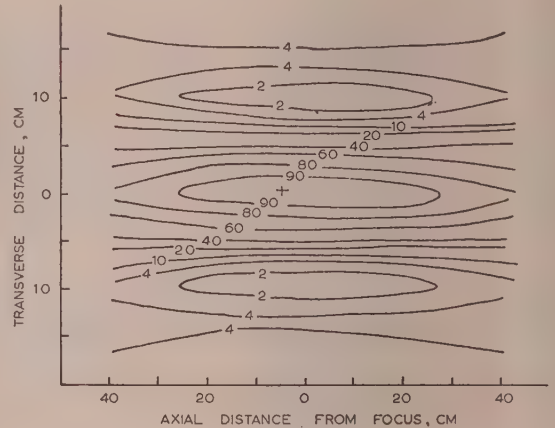


Fig. 4.—Theoretical intensity contours in x or y axial planes.

$$\begin{aligned} \eta_0 &= \xi_0 = 60 \text{ cm} \\ \xi_0 &= 372 \text{ cm} \\ \lambda_0 &= 3.20 \text{ cm} \end{aligned}$$

which will be compared with this diagram in a later Section. The main feature is the long, narrow, approximately ellipsoidal region of high intensity near the focal point. 'Side lobes' in the transverse plane will be noticed at about 15 cm from the axis. Further simplification is possible if attention is confined to the field in the focal plane or on the axis.

In the focal plane $z = 0$, and eqn. (21) can be integrated easily. The result is

$$\phi = \frac{4j\eta_0\xi_0}{\lambda\xi_0^2} \frac{\sin p}{p} \frac{\sin p'}{p'} \quad (24)$$

When $x \ll \xi_0$ we can write $\sin \theta \approx x/\xi_0$, where θ is the angular deviation off the axis in the xz -plane.

$$\text{Hence,} \quad \frac{\sin p}{p} \approx \frac{\sin(k\eta_0 \sin \theta)}{k\eta_0 \sin \theta} \quad (25)$$

which will be recognized as the radiation pattern of a uniformly illuminated equiphase aperture distribution. Expressed in this form, the pattern is independent of the focal distance ξ_0 .

This familiar radiation pattern can be regarded as a special case of the field distribution in the focal plane of a lens system when the lens is focused at infinity. The problem under discussion is more general because distances greater than the focal distance must be considered in connection with the variation of the field along the axis; this case cannot, of course, arise if the lens is focused at infinity.

Eqn. (24) is plotted against p to show the transverse field pattern in the focal plane, in Fig. 7.

Consider now the field variation along the axis, with $x = y = 0$.

In this case* eqn. (21) can be expressed in terms of the Fresnel integrals $C(x)$ and $S(x)$ defined by

$$C(x) + jS(x) = \int_0^x e^{j\pi t^2/2} dt \quad (26)$$

Attention will be confined to a square aperture in order to simplify the subsequent formulae, i.e. let $\xi_0 = \eta_0$.

The result for q positive is

$$\phi = \frac{4j\eta_0^2}{\lambda\xi_0^2} \exp(-jkz) \frac{\pi}{q} \left[C\left(\sqrt{\frac{q}{\pi}}\right) + jS\left(\sqrt{\frac{q}{\pi}}\right) \right]^2 \quad (27)$$

For q negative, the expression becomes

$$\phi = \frac{4j\eta_0^2}{\lambda\xi_0^2} \exp(-jkz) \frac{\pi}{|q|} \left[C\left(\sqrt{\frac{|q|}{\pi}}\right) - jS\left(\sqrt{\frac{|q|}{\pi}}\right) \right]^2 \quad (28)$$

The factors in square brackets in (27) and (28) have equal moduli but opposite phase angles for given values of $|q|$.

The variation of phase along the axis can be expressed as the sum of a linear phase change due to the term $\exp(-jkz)$, together with an additional phase change due to the bracketed factor. The resultant phase is given by

$$\theta = -kz \pm \arctan \left\{ \frac{S\left(\sqrt{\frac{|q|}{\pi}}\right)}{C\left(\sqrt{\frac{|q|}{\pi}}\right)} \right\} \quad (29)$$

where $q = kz\eta_0^2/\xi_0^2$; the upper sign is to be taken for q positive, the lower sign for q negative.

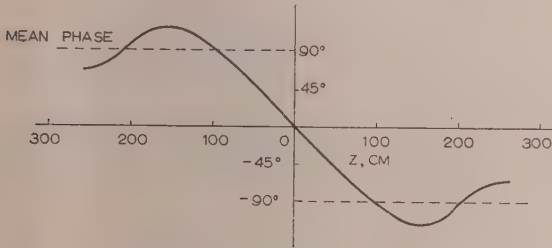


Fig. 5.—Variation in θ_2 with z .

$$\begin{aligned} \eta_0 &= \xi_0 = 60 \text{ cm} \\ \xi_0 &= 372 \text{ cm} \\ \lambda_0 &= 3.20 \text{ cm} \end{aligned}$$

In Fig. 5 the last term in eqn. (29) is plotted as a function of z for a wavelength of 3.2 cm. It will be seen that there is a change of 180° due to this term as z varies from $-\infty$ to $+\infty$. This is the phase change referred to in Section 1. It will be noticed that the major part of the change of phase occurs relatively slowly, over a distance of about one hundred wavelengths, and that the phase varies linearly with distance for small distances from the focal plane ($z = 0$).

If an approximation be made in eqn. (26) by writing $e^{j(\pi/2)t^2} \approx 1 + j(\pi/2)t^2$, small-argument approximations are obtained for $C(x)$ and $S(x)$, which, when substituted in eqn. (29), yield

$$\theta \approx -kz \left(1 - \frac{\eta_0^2}{3\xi_0^2} \right)$$

The axial wavelength λ_a is therefore greater than the free-space wavelength λ , the relation being

$$\lambda_a \approx \lambda \left(1 + \frac{\eta_0^2}{3\xi_0^2} \right) \quad (30)$$

* Eqn. (21) can also in general be expressed in terms of Fresnel integrals. Lommel's formulation is here used because his solution was the first historically.

This increase in axial wavelength near the focus can be interpreted physically in the following way. The field near the focus can be regarded as the sum of an infinite number of plane waves. Each plane wave is associated with a ray through the focus which makes an angle α with the axis such that $0 \leq \alpha \leq \arctan(\sqrt{2}\eta_0/\xi_0)$. The axial wavelength associated with a plane wave travelling at an angle α to the axis is $\lambda \sec \alpha$, where λ is the wavelength in the direction of propagation of the wave. This point of view is often adopted in explaining the increased axial wavelength λ_g in a rectangular waveguide for the H_{01} mode, which can be regarded as the sum of two plane waves.

The axial wavelengths associated with the infinite set of plane waves range from λ to $\lambda \sec [\arctan(\sqrt{2}\eta_0/\xi_0)]$, or, to a first approximation, from λ to $\lambda \left(1 + \frac{\eta_0^2}{\xi_0^2} \right)$.

The resultant axial wavelength must lie somewhere between these limits, and from this point of view eqn. (30) is physically reasonable.

The variation of amplitude with z , obtained from eqn. (28) is plotted in Fig. 6 for the dimension of aperture and focal distance used in the experiment.

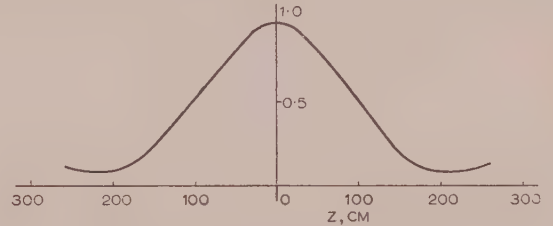


Fig. 6.—Variation in intensity on axis.

$$\begin{aligned} \eta_0 &= \xi_0 = 60 \text{ cm} \\ \xi_0 &= 372 \text{ cm} \\ \lambda_0 &= 3.20 \text{ cm} \end{aligned}$$

(2.4.2) Tapered Aperture Distribution.

Assume that the aperture distribution is parabolically tapered; then

$$A(\eta, \xi) = \left(1 - a \frac{\eta^2}{\eta_0^2} \right) \left(1 - a \frac{\xi^2}{\xi_0^2} \right) \quad (31)$$

With this functional form the integrations in eqn. (11) are again independent. The final result for ϕ is

$$\phi = \frac{j\eta_0\xi_0}{\lambda\xi_0^2} \exp(-jkz) I_2 I_2' \quad (32)$$

$$\text{where } I_2 = I_1 \left(1 + j \frac{a}{q} - a \frac{p^2}{q^2} \right)$$

$$+ a \sqrt{\left(\frac{2\pi}{q} \right)} \exp \left(-j \frac{q}{2} \right) \left[J_{1/2}(p) \left(\frac{p^2}{q^2} - \frac{j}{q} \right) + j J_{3/2}(p) \frac{p}{q} \right] \quad (33)$$

$$\text{and } I_1 = \sqrt{\left(\frac{2\pi}{q} \right)} \exp \left(-j \frac{q}{2} \right) [U_{1/2}(q, p) + j U_{3/2}(q, p)] \quad (34)$$

with analogous expressions for I_2' and I_1' .

Simpler formulae can be obtained for the field in the focal plane and on the axis.

In the focal plane,

$$\phi = \frac{4j\eta_0\xi_0}{\lambda\xi_0^2} F(p)F(p') \quad (35)$$

$$\text{where } F(p) = \frac{\sin p}{p} \left(1 - a + \frac{2a}{p^2} \right) - \frac{4a \cos p}{p^2} \quad (36)$$

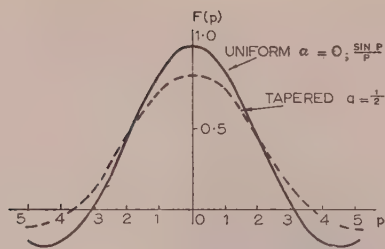


Fig. 7.—Effect of tapered amplitude in aperture on the focal-plane amplitude.

Taper is of the form $(1 - a\frac{\eta^2}{\eta_0^2})$.

In Fig. 7 the function $F(p)$ is plotted against p for $a = \frac{1}{2}$, and, for comparison for $a = 0$, when $F(p)$ reduces to the distribution $\sin p/p$ associated with a uniform aperture distribution.

The taper has the effect, well known in aerial design, of broadening the main lobe and diminishing the relative intensity of the side lobes.

The effect of the taper on the axial wavelength can be calculated directly from eqn. (11) by using the small-argument approximation for $\exp [j(k\eta^2/2\zeta_0^2)z]$.¹ The result is

$$\lambda_a = \lambda \left[1 + \frac{\eta_0^2}{3\zeta_0^2} \left(1 - \frac{3}{2}a \right) \right] \quad (37)$$

It will be observed that this expression reduces to eqn. (30), as it should, when $a = 0$.

(3) EXPERIMENTAL PROCEDURE AND RESULTS

The theoretical results obtained in Section 2 refer to rather idealized field patterns. The experimental procedure was designed to produce fields as near to these theoretical patterns as could be achieved with the apparatus available.

The field patterns were measured by means of the spinning-dipole technique described by Cullen and Parr.¹² The general arrangement of the apparatus is shown in Fig. 8. The lens used

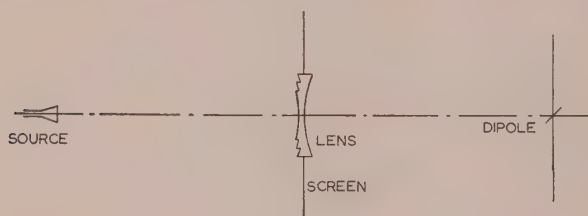


Fig. 8.—Schematic layout of apparatus.

was a three-step 'egg-box' type with a square aperture 120×120 cm. It is two-point corrected for scanning, and was designed to produce a parallel beam. Its focal length is 192 cm.

The transmitting horn was situated at a point twice the focal length from one side of the lens, and the field was measured in a region approximately the same distance from the other side of the lens. The position of the transmitting horn was adjusted to give a symmetrical pattern in a plane transverse to the direction of propagation near the focus and to give as good an approximation to a spherical wavefront as possible.

The vertical component of the electric field was measured in the horizontal median plane.

The amplitude distribution across the lens aperture was measured with a simple crystal probe and is shown in Fig. 9. An average curve drawn through these points gives a taper factor, a , of about 0.4. The scatter of these points is due to the egg-box

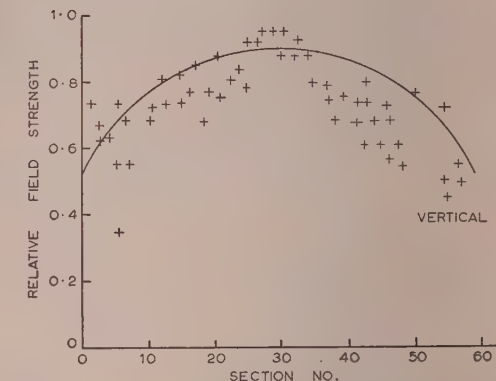
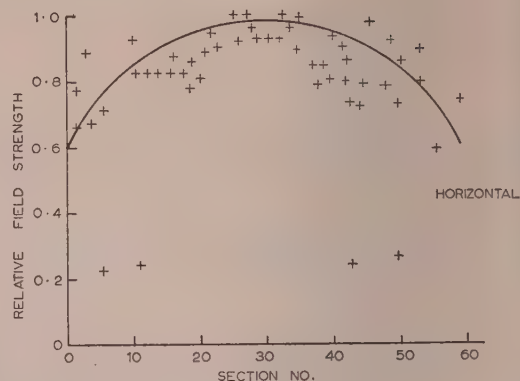


Fig. 9.—Aperture distribution measured at the lens surface.

structure of the lens and the steps in its profile. The focal-plane intensity calculated from this value of taper is shown in Fig. 10, together with the experimental points measured in the focal

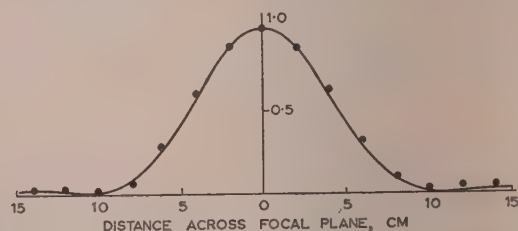


Fig. 10.—Focal-plane intensity calculated for $a = 0.4$. Experimental observations are also shown.

plane. It will be seen that there is close agreement between the experimental and theoretical points. The agreement is surprisingly close when the large fluctuations in amplitude across the aperture are borne in mind.

This insensitiveness of the focal-plane pattern to wide variations in the aperture amplitude distribution is closely related to the more familiar insensitiveness of radiation patterns of aerials to amplitude variation across the radiating aperture.

The field pattern in the region around the focus was measured, and is shown in Fig. 11, which is drawn in contours of equal intensity and lines of equal phase. The intensity contours are mostly of the form to be expected. Some of the distortion may be due to the rough amplitude distribution across the lens aperture and some to the presence of other reflecting objects near the apparatus.

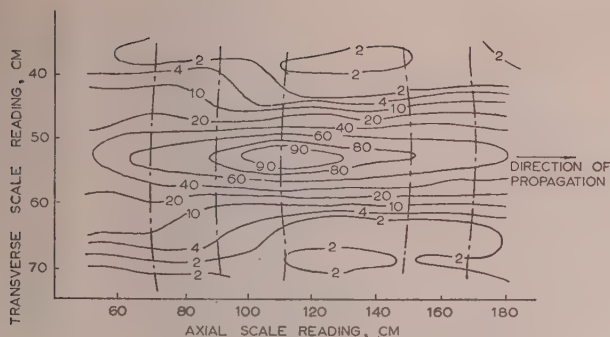


Fig. 11.—Contours of equal intensity and equal phase.

It will be noticed that the deviations from the general form of the theoretical contours are most marked in regions of very low field strength, where unwanted reflections are more serious and accuracy of measurement is less. Detailed agreement cannot be expected, because the theoretical contours are drawn for a uniform aperture illumination. The labour of computation for the tapered illumination is very much greater than for uniform illumination, and only focal-plane and axial distributions have been studied in detail for this case.

The lines of constant phase in Fig. 11 show the change from incoming to outgoing waves on passing the focus. This change is closely associated with the phase change of π through the focus referred to earlier. Owing to the changing curvature of successive wavefronts, the wavelength along the axis is greater than the free-space wavelength. This axial wavelength has been calculated from eqn. (37) for $\lambda = 3.200$ cm and $a = 0.4$, giving $\lambda_a = 3.224$ cm. This wavelength was measured by observing the distance through which the dipole must be moved to produce a phase change equivalent to 50 wavelengths. The measured value of λ_a is 3.23 ± 0.01 cm, which is in satisfactory agreement with the theory. The effect of the taper is very small, the theoretical value of λ_a for a uniform aperture distribution being 3.228 cm. The so-called 'anomalous phase change' at the focus is, in fact, a smooth change in phase with distance and not a sudden jump of 180° . Near the focus it is apparent only in the changed axial wavelength to which reference has been made.

In the transverse plane there is a sharp change in phase near the minima of intensity. In the focal plane this phase change is 180° . It cannot be measured directly with the apparatus used, but may be shown to occur in the following way.

The dipole is first set at a point A on the axis on one side of the focus as shown in Fig. 12, and the phase shifter is adjusted to make the galvanometer reading a maximum. The dipole is then moved along the axis to a point B on the other side of the focus. The number of successive maxima of the galvanometer deflection is noted. The dipole is then moved out along a line of constant phase to a point C in the first side-lobe. It is then moved back along a line CPD parallel to the axis through the same number of maxima as before, and is finally brought into the axis along a line of constant phase. It is then found that the final position, E, of the dipole is further from the focus than the first position by one wavelength. Thus, the phase change along the straight line EB is 360° greater than the phase change along the straight line DC. By symmetry, the phase change EO is 180° greater than the phase change DP. But the field at E is in phase with the field at D, since both points lie on an equiphase surface; therefore the field at P is 180° out of phase with the field at O. Hence, the theoretically predicted phase change is verified.

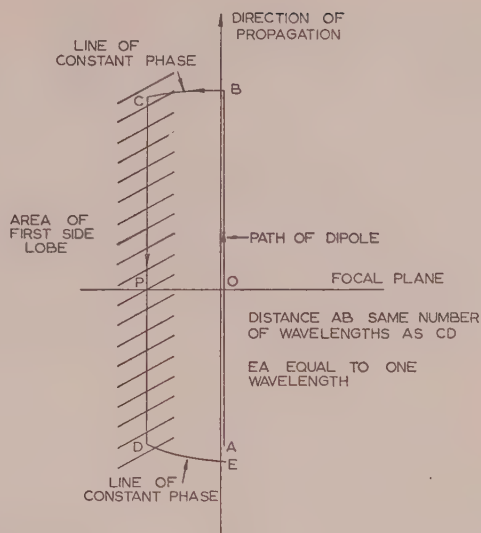


Fig. 12.—Diagram of path of dipole, demonstrating phase change in the focal plane between main and side lobes.

(4) CONCLUSIONS

The theory of the field distribution near the focal point of a microwave lens has been considered in some detail. Previous work on the problem based on scalar wave-diffraction theory has been extended to allow for a tapered aperture illumination, and the application of this scalar theory to the calculation of the principal component of the electric-field vector in the microwave problem has been justified.

Experimental work has confirmed the theory within the limitation of the available equipment. Particularly satisfactory agreement has been found with the theoretically predicted transverse field pattern in the focal plane, and with the theoretically predicted increase in axial wavelength in the vicinity of the focus. The latter effect is related to the so-called 'anomalous phase change' of 180° on passing through the focus. A phase change of 180° between main lobe and first side-lobe in the focal plane which the theory predicts has also been verified experimentally.

(5) ACKNOWLEDGMENTS

The authors wish to thank Professor H. E. M. Barlow for his interest in the work described and for providing laboratory facilities. They are particularly grateful to Dr. K. Milne and the British Thomson-Houston Co., Ltd., for the loan of the lens and for providing detailed information on its design, and they also acknowledge the helpful suggestions and criticism in the preparation of the paper given by Dr. J. Brown.

One of the authors (P. A. Matthews) is indebted to the Department of Scientific and Industrial Research for a grant which made it possible for him to undertake the work.

(6) REFERENCES

- (1) ATRY, G. B.: 'On the Diffraction of an Object Glass with a Circular Aperture', *Transactions of the Cambridge Philosophical Society*, 1834, **5**, p. 283.
- (2) LOMMEL, E.: 'Die Beugungserscheinungen einer Öffnung und eines kreisrunden Schirmchens', *Königlich Bayerischen Akademie der Wissenschaften, Abhandlung der Mathematisch-Physikalischen Klasse*, 1886, **15**, p. 229.
'Die Beugungserscheinungen geradlinig begrenzter Schirme', *ibid.*, p. 529.

- (3) GOUY, A.: 'Sur une propriété nouvelle des ondes lumineuses', *Comptes Rendus de Paris*, 1890, **110**, p. 1251.
- (4) JAUBIN, P.: 'Sur le passage d'une onde par un foyer', *Comptes Rendus de Paris*, 1892, **115**, p. 932.
- (5) ZEEMAN, P.: 'Ein Experiment über die sogenannte anomale Fortpflanzung von Wellen', *Zeitschrift für Physik*, 1900, **1**, p. 542.
- (6) REICHE, F.: 'Über die anomale Fortpflanzung von Kugelnwellen beim Durchgang durch Brennpunkte', *Annalen der Physik*, 1909, **29**, pp. 65 and 401.
- (7) DEBYE, F.: 'Das Verhalten von Lichtwellen in der Nähe eines Brennpunktes oder einer Brennnlinie', *Annalen der Physik*, 1909, **30**, p. 755.
- (8) STRATTON, F.: 'Electromagnetic Theory' (McGraw-Hill, 1941).
- (9) SOMMERFELD, A.: 'Optics' (Academic Press, New York, 1954).
- (10) SCHELKUNOFF, S. A.: 'Symposium on the Theory of Electromagnetic Waves' (Interscience Publishers, New York, 1951).
- (11) WATSON, G. N.: 'Theory of Bessel Functions' (Cambridge University Press, 1944).
- (12) Tables of Spherical Bessel Functions (Columbia University Press, 1947).
- (13) SILVER, S.: 'Microwave Antenna Theory and Practice' (McGraw-Hill, 1949).
- (14) CULLEN, A. L., and PARR, J. C.: 'A New Perturbation Method for Measuring Microwave Fields in Free Space', *Proceedings I.E.E.*, Paper No. 1921 R, November, 1955 (**102 B**, p. 836).

INDEX TO VOLUME 103, PART C

1956

ABBREVIATIONS

(P)—Paper.

(D)—Discussion.

A

- A.C.-controlled transducers. (*See* Transducers.)
 — network analyser. (*See* Network.)
 Adcock direction-finder. (*See* Direction-finder.)
 ADKINS, B., and SEN, S. K. (*See* SEN.)
 AKED, A. Peak-voltage measurements of standard impulse voltage waves. (P), 186.
 ALLEN, C. P., and CLARKE, G. M. Interpretation of wavelength measurements on tape helices. (P), 171.
 ALLISON, J., and BENSON, F. A. Attenuation and permeability of ferromagnetic waveguides between 9 000 and 9 675 Mc/s. (P), 205.
 ALTSCHULER, H. M., and OLINER, A. A. Microwave measurements with lossy variable termination. (P), 392.
 Amplitude response, correlation between decay time and. S. DEMCZYNSKI, (P), 64.
 Analogue computing, use of function generators in. E. G. C. BURT and O. H. LANGE, (P), 51.
 Anisotropic waveguides. (*See* Waveguides.)
 ARNOLD, A. H. M., and HILL, J. J. Audio-frequency dynamometer wattmeter. (P), 325.
 Arrays, broadside, determination of true side-lobe level of. R. H. T. BATES and J. ELLIOTT, (P), 307.
 ASPDEN, H.
 Eddy-current anomaly in electrical sheet steel. (P), 272.
 Magnetic time-lag effects in solid steel cores. (P), 279.
 Attenuation and permeability of ferromagnetic waveguides between 9 000 and 9 675 Mc/s. J. ALLISON and F. A. BENSON, (P), 205.
 Audio-frequency dynamometer wattmeter. A. H. M. ARNOLD and J. J. HILL, (P), 325.

B

- BAILEY, R. A. Resonant-cavity torque-operated wattmeter for microwave power. (P), 59.
 BAIN, W. C. Possible errors of a particular wide-aperture direction-finder. (P), 313.
 BAIN, W. C., and GLASS, C. B.-I. Polarization of very long radio waves reflected from the ionosphere at oblique incidence. (P), 447.
 Band 8-9 mm, reflex klystron oscillator for. D. J. WOOTTON and A. F. PEARCE, (P), 104.
 BANKS, J. H., and LEWIS, W. E. (*See* LEWIS.)
 BANKS, J. H., and PARTON, K. C. Use of Blackburn a.c. network analyser in analysis of power system faults. (P), 342.
 BARRETT, J. F., and COALES, J. F. Analysis of non-linear control systems with random inputs. (P), 190.
 BATES, R. H. T., and ELLIOTT, J. Determination of true side-lobe level of long broadside arrays from radiation-pattern measurements made in the Fresnel region. (P), 307.
 Bearing variations, h.f., on Adcock direction-finder. E. N. BRAMLEY, (P), 350.
 BENSON, F. A., and ALLISON, J. (*See* ALLISON.)
 BIRCH, J., FRITH, A. G., FERGUSON, A. C. L., MILES, R. H. A., and WERNER, J. F. Use of ethylene diamine tartrate for piezo-electric filter elements. (P), 420.
 Blackburn a.c. network analyser. (*See* Network.)
 BRAMLEY, E. N. H.F. bearing variations on Adcock direction-finder. (P), 350.
 BROADBENT, D. Steady-state stability of synchronous generators. (D), 231.

VOL. 103, PART C.

- BROWN, J., and COLLIN, R. E. (*See* COLLIN.)
 BRUCK, R. W., and MESSERLE, H. K. Steady-state stability of synchronous generators as affected by regulators and governors. (P), 24; (D), 232.
 BURT, E. G. C., and LANGE, O. H. Function generators based on linear interpolation with applications to analogue computing. (P), 51.

C

- CARTER, G. W. Surface loss in laminated pole-face. (D), 231.
 CEDERBAUM, I.
 Analysis of linear n -port networks. (P), 267.
 Properties of transfer function of unbalanced RC networks. (P), 400.
 Chebyshev patterns, Fourier coefficients for. H. E. SALZER, (P), 286.
 Circuit (equivalent) of axially unsymmetrical waveguide junction, calculation of. R. E. COLLINS and J. BROWN, (P), 121.
 —, RC, suppression of switching transients by. (D), 233.
 CLARKE, G. M., and ALLEN, C. P. (*See* ALLEN.)
 COALES, J. F., and BARRETT, J. F. (*See* BARRETT.)
 Coaxial resistor mounts. (*See* Resistor.)
 Coefficients, Fourier, for Chebyshev patterns. H. E. SALZER, (P), 286.
 Coils, iron-cored, inductance of. P. HAMMOND, (P), 249.
 COLLIN, R. E., and BROWN, J.
 Calculation of equivalent circuit of axially unsymmetrical waveguide junction. (P), 121.
 Design of quarter-wave matching layers for dielectric surfaces. (P), 153.
 Control systems, non-linear, with random inputs, analysis of. J. F. BARRETT and J. F. COALES, (P), 190.
 Correlation between decay time and amplitude response. S. DEMCZYNSKI, (P), 64.
 COVENTRY, A. F. Resultant reactive power of overhead lines. (P), 334.
 CULLEN, A. L., and MATTHEWS, P. A. (*See* MATTHEWS.)
 CULLWICK, E. G.
 Electromagnetic momentum and electron inertia in a current circuit. (P), 159.
 Magnetic energy and electron inertia in a superconducting sphere. (P), 441.
 Current circuit, electromagnetic momentum and electron inertia in. E. G. CULLWICK, (P), 159.
 Cylinders, conducting, radiation patterns of circumferential slots on. J. R. WAIT and J. KATES, (P), 289.

D

- DEARDS, S. R. Matrix methods for evaluation of simultaneous faults in three-phase systems. (D), 233.
 Decay time and amplitude response, correlation between. S. DEMCZYNSKI, (P), 64.
 DEMCZYNSKI, S. Correlation between decay time and amplitude response. (P), 64.
 Dielectric surfaces, quarter-wave matching layers for. R. E. COLLIN and J. BROWN, (P), 153.
 Diffraction of electromagnetic wave. (*See* Wave.)
 Direction-finder, Adcock, h.f. bearing variations on. E. N. BRAMLEY, (P), 350.
 —, wide-aperture, possible errors of. W. C. BAIN, (P), 313.
 Discrimination of synchronized oscillator against interfering tones and noise. D. G. TUCKER and G. G. JAMIESON, (P), 129.
 Dynamometer wattmeter, audio-frequency. A. H. M. ARNOLD and J. J. HILL, (P), 325.

E

- Eddy-current anomaly in electrical sheet steel. H. ASPDEN, (P), 272.
 EL-DINE, M. E. Z., and TROPPER, H. Electric strength of transformer oil. (P), 35.
 Electric strength. (*See Strength.*)
 Electrical machines. (*See Machines.*)
 — sheet steel. (*See Steel.*)
 Electromagnetic momentum and electron inertia in a current circuit. E. G. CULLWICK, (P), 159.
 — wave. (*See Wave.*)
 ELLIOTT, J., and BATES, R. H. T. (*See BATES.*)
 Energy, magnetic, and electron inertia in a superconducting sphere. E. G. CULLWICK, (P), 441.
 Errors of a particular wide-aperture direction-finder. W. C. BAIN, (P), 313.
 Ethylene diamine tartrate, use of, for piezo-electric filter elements. J. BIRCH, A. G. FRITH, A. C. L. FERGUSON, R. H. A. MILES and J. F. WERNER, (P), 420.

F

- FARMER, E. D. Junction admittance between waveguides of arbitrary cross-sections. (P), 145.
 Faults, power system. (*See Power.*)
 —, simultaneous, in three-phase systems, evaluation of. (D), 233.
 Feedback-system analogue based on discontinuous delay-line synthesizer, properties of. R. M. F. HOUTAPPEL, (P), 367.
 FELLOWS, G. E., and MIDDLETON, D. Experimental study of intensity spectra after half-wave rectification of signals in noise. (P), 243.
 FERGUSON, A. C. L., MILES, R. H. A., WERNER, J. F., BIRCH, J., and FRITH, A. G. (*See BIRCH.*)
 Ferromagnetic waveguides. (*See Waveguides.*)
 Field distribution at axial focus of square microwave lens. P. A. MATTHEWS and A. L. CULLEN, (P), 449.
 Fields, third-harmonic and zero-sequence, theory of. G. H. RAWCLIFFE and B. C. McDERMOTT, (P), 212.
 Filament structures (transmitting-valve), mutual heating in. W. J. POHL, (P), 224.
 Filter elements, piezo-electric, use of ethylene diamine tartrate for. J. BIRCH, A. G. FRITH, A. C. L. FERGUSON, R. H. A. MILES and J. F. WERNER, (P), 420.
 FINLAY, E. A. Suppression of switching transients by shunt RC circuit. (D), 233.
 Fourier coefficients. (*See Coefficients.*)
 Frequency band 0-4000 Mc/s, coaxial resistor mounts for. I. A. HARRIS, (P), 1.
 Frequency-response method, application of, to electrical machines. S. K. SEN and B. ADKINS, (P), 378.
 Fresnel region, radiation-pattern measurements made in. (*See Radiation.*)
 FRITH, A. G., FERGUSON, A. C. L., MILES, R. H. A., WERNER, J. F., and BIRCH, J. (*See BIRCH.*)
 Function generators. (*See Generators.*)

G

- Generators (function) based on linear interpolation with applications to analogue computing. E. G. C. BURT and O. H. LANGE, (P), 51.
 —, synchronous, relative dynamic stability of. H. K. MESSERLE, (P), 234.
 —, synchronous, steady-state stability of, as affected by regulators and governors. H. K. MESSERLE and R. W. BRUCK, (P), 24; (D), 231.
 GLASS, C. B.-I., and BAIN, W. C. (*See BAIN.*)
 GOOD, I. J. Terminology and notation in information theory. (P), 200.

H

- HAMMOND, P.
 Inductance of iron-cored coils. (P), 249.
 Magnetic screening effect of iron tubes. (P), 112.
 HARRIS, I. A. Theory and design of coaxial resistor mounts for frequency band 0-4000 Mc/s. (P), 1.
 HEAD, J. W., and WILSON, W. P. Laguerre functions: tables and properties. (P), 428.

- Heating, mutual, in transmitting-valve filament structures. W. J. POHL, (P), 224.
 H.F. bearing variations. (*See Bearing.*)
 HILL, J. J., and ARNOLD, A. H. M. (*See ARNOLD.*)
 HOUTAPPEL, R. M. F. Properties of feedback-system analogue based on discontinuous delay-line synthesizer. (P), 367.

I

- Impulse voltages waves. (*See Waves.*)
 Inductance of iron-cored coils. P. HAMMOND, (P), 249.
 Information theory, terminology and notation in. I. J. GOOD, (P), 200.
 Inputs, random, non-linear control systems with. J. F. BARRETT and J. F. COALES, (P), 190.
 Integral control with torque limitation. J. C. WEST and M. J. SOMERVILLE, (P), 407.
 Intensity spectra, study of, after half-wave rectification of signals in noise. G. E. FELLOWS and D. MIDDLETON, (P), 243.
 Interfering tones and noise. (*See Tones.*)
 Ionosphere, polarization of very long radio waves reflected from. W. C. BAIN and C. B.-I. GLASS, (P), 447.
 Iron-cored coils, inductance of. P. HAMMOND, (P), 249.
 Iron tubes, magnetic screening effect of. P. HAMMOND, (P), 112.

J

- JAMIESON, G. G., and TUCKER, D. G. (*See TUCKER.*)
 Junction admittance between waveguides of arbitrary cross-sections. E. D. FARMER, (P), 145.

K

- KARBOWIAK, A. E. Microwave propagation in anisotropic waveguides. (P), 139.
 KATES, J., and WAIT, J. R. (*See WAIT.*)
 KRABBE, U. Residual time-constant of self-saturating (auto-excited) transducers. (P), 71.

L

- Laguerre functions: tables and properties. J. W. HEAD and W. P. WILSON, (P), 428.
 — polynomials, table of. L. J. SLATER, (P), 46
 LANGE, O. H., and BURT, E. G. C. (*See BURT.*)
 LAW, T. S., and MILNES, A. G. (*See MILNES.*)
 LEWIS, W. E., and BANKS, J. H. Matrix methods for evaluation of simultaneous faults in three-phase systems. (D), 233.
 Linear interpolation, function generators based on. E. G. C. BURT and O. H. LANGE, (P), 51.
 — *n*-port networks. (*See Networks.*)
 — servo mechanism. (*See Servo.*)
 Lossy variable termination, microwave measurements with. H. M. ALTSCHULER and A. A. OLINER, (P), 392.

M

- McDERMOTT, B. C., and RAWCLIFFE, G. H. (*See RAWCLIFFE.*)
 Machines, electrical, application of frequency-response method to. S. K. SEN and B. ADKINS, (P), 378.
 Magnetic energy. (*See Energy.*)
 — screening. (*See Screening.*)
 — time-lag effects. (*See Time-lag.*)
 Magnetron, millimetre-wave. J. R. M. VAUGHAN, (P), 95.
 Magnetrons using spatial-harmonic operation, properties of. R. G. ROBERTSHAW and W. E. WILLSHAW, (P), 297.
 Matching layers, quarter-wave, for dielectric surfaces, design of. R. E. COLLIN and J. BROWN, (P), 153.
 Matrix methods for evaluation of simultaneous faults in three-phase systems. (D), 233.
 MATTHEWS, P. A., and CULLEN, A. L. Study of field distribution at axial focus of square microwave lens. (P), 449.
 Measurements, microwave. H. M. ALTSCHULER and A. A. OLINER, (P), 392.
 —, peak-voltage, of standard impulse voltage waves. A. AKED, (P), 186.
 —, wavelength, on tape helices, interpretation of. C. P. ALLEN and G. M. CLARKE, (P), 171.

- MESSERLE, H. K. Relative dynamic stability of large synchronous generators. (P), 234.
- MESSERLE, H. K., and BRUCK, R. W. Steady-state stability of synchronous generators as affected by regulators and governors. (P), 24; (D), 232.
- Method for finding 'best linear servo mechanism'. H. H. ROSENBRICK, (P), 260.
- Microwave lens, square, field distribution at axial focus of. P. A. MATTHEWS and A. L. CULLEN, (P), 449.
- measurements with a lossy variable termination. H. M. ALTSCHULER and A. A. OLINER, (P), 392.
- power, resonant-cavity torque-operated wattmeter for. R. A. BAILEY, (P), 59.
- propagation in anisotropic waveguides. A. E. KARBOWIAK, (P), 139.
- MIDDLETON, D., and FELLOWS, G. E. (See FELLOWS.)
- MILES, R. H. A., WERNER, J. F., BIRCH, J., FRITH, A. G., and FERGUSON, A. C. L. (See BIRCH.)
- MILLAR, R. F. Diffraction of electromagnetic wave by aperture in plane screen. (P), 177.
- Millimetre-wave magnetron. J. R. M. VAUGHAN, (P), 95.
- MILNES, A. G., and LAW, T. S. A.C.-controlled transducers. (P), 81.
- Motor, Schrage, operating at synchronous speed, second-order torque components in. I. THOMAS, (P), 11.
- MUKHERJI, K. C. Surface loss in laminated pole-face. (D), 230.
- Mutual heating. (See Heating.)

N

- n -port networks. (See Networks.)
- Network analyser (a.c.), Blackburn, for use in analysis of power system faults. J. H. BANKS and K. C. PARTON, (P), 342.
- Networks, n -port, analysis of. I. CEDERBAUM, (P), 267.
- , RC, properties of transfer function of. I. CEDERBAUM, (P), 400.
- NEVILLE, S. Reluctance of teeth of slotted armature. (P), 338.
- Noise, discrimination of synchronized oscillator against. D. G. TUCKER and G. G. JAMIESON, (P), 129.
- , rectification of signals in. G. E. FELLOWS and D. MIDDLETON, (P), 243.
- Non-linear control systems. (See Control.)
- Notation in information theory. (See Information.)

O

- Oil (transformer), electric strength of. M. E. Z. EL-DINE and H. TROPPER, (P), 35.
- OLINER, A. A., and ALTSCHULER, H. M. (See ALTSCHULER.)
- Oscillator, reflex klystron, for 8-9 mm band. D. J. WOOTTON and A. F. PEARCE, (P), 104.
- , synchronized, discrimination of, against interfering tones and noise. D. G. TUCKER and G. G. JAMIESON, (P), 129.
- Overhead lines, resultant reactive power of. A. F. COVENTRY, (P), 334.

P

- PARTON, K. C., and BANKS, J. H. (See BANKS.)
- Peak-voltage measurements of standard impulse voltage waves. A. AKED, (P), 186.
- PEARCE, A. F., and WOOTTON, D. J. (See WOOTTON.)
- Permeability and attenuation of ferromagnetic waveguides. (See Waveguides.)
- Piezo-electric filter elements. (See Filter.)
- Plane screen, diffraction of electromagnetic wave by aperture in. R. F. MILLAR, (P), 177.
- POHL, W. J. Mutual heating in transmitting-valve filament structures. (P), 224.
- Polarization of very long radio waves. W. C. BAIN and C. B.-I. GLASS, (P), 447.
- Pole-face, laminated, surface loss in. (D), 230.
- Polynomials, Laguerre, table of. L. J. SLATER, (P), 46.
- Power, reactive, of overhead lines. A. F. COVENTRY, (P), 334.
- system faults, use of Blackburn a.c. network analyser for. J. H. BANKS and K. C. PARTON, (P), 342.

Q

- Quarter-wave matching layers for dielectric surfaces, design of. R. E. COLLIN and J. BROWN, (P), 153.

R

- Radiation-pattern measurements made in the Fresnel region to determine true side-lobe level of long broadside arrays. R. H. T. BATES and J. ELLIOTT, (P), 307.
- Radiation patterns of circumferential slots on conducting cylinders. J. R. WAIT and J. KATES, (P), 289.
- Radio waves. (See Waves.)
- Random inputs. (See Inputs.)
- RAWCLIFFE, G. H., and McDERMOTT, B. C. Theory of third-harmonic and zero-sequence fields. (P), 212.
- RC circuit. (See Circuit.)
- networks. (See Networks.)
- Reactive power of overhead lines. (See Power.)
- READ, J. C. Suppression of switching transients by shunt RC circuit. (D), 233.
- Rectification of signals in noise. (See Noise.)
- Reflex klystron oscillator for 8-9 mm band. D. J. WOOTTON and A. F. PEARCE, (P), 104.
- Regulators and governors, steady-state stability of synchronous generators as affected by. H. K. MESSERLE and R. W. BRUCK, (P), 24; (D), 231.
- Reluctance of teeth of slotted armature. S. NEVILLE, (P), 338.
- Resistor mounts, coaxial, for frequency band 0-4000 Mc/s. I. A. HARRIS, (P), 1.
- Resonant-cavity torque-operated wattmeter for microwave power. R. A. BAILEY, (P), 59.
- Resultant reactive power of overhead lines. A. F. COVENTRY, (P), 334.
- ROBERTSHAW, R. G., and WILLSHAW, W. E. Properties of magnetrons using spatial-harmonic operation. (P), 297.
- ROSENBRICK, H. H. Approximate method for finding 'best linear servo mechanism'. (P), 260.

S

- SALZER, H. E. Note on the Fourier coefficients for Chebyshev patterns. H. E. SALZER, (P), 286.
- Schrage motor. (See Motor.)
- Screening (magnetic) effect of iron tubes. P. HAMMOND, (P), 112.
- SEN, S. K., and ADKINS, B. Application of frequency-response method to electrical machines. (P), 378.
- Servo mechanism, linear, approximate method for finding the best. H. H. ROSENBRICK, (P), 260.
- Sheet steel. (See Steel.)
- Side-lobe level of long broadside arrays, determination of. R. H. T. BATES and J. ELLIOTT, (P), 307.
- Signals, half-wave rectification of, in noise. G. E. FELLOWS and D. MIDDLETON, (P), 243.
- SLATER, LUCY J. Table of Laguerre polynomials. (P), 46.
- Slots, circumferential, on moderately large conducting cylinders, radiation patterns of. J. R. WAIT and J. KATES, (P), 289.
- Slotted armature, reluctance of teeth of. S. NEVILLE, (P), 338.
- SOMERVILLE, M. J., and WEST, J. C. (See WEST.)
- Spatial-harmonic operation, properties of magnetrons using. R. G. ROBERTSHAW and W. E. WILLSHAW, (P), 297.
- Speed, synchronous, second-order torque components in Schrage motor operating at. I. THOMAS, (P), 11.
- Stability, dynamic, of large synchronous generators. H. K. MESSERLE, (P), 234.
- , steady-state, of synchronous generators. H. K. MESSERLE and R. W. BRUCK, (P), 24; (D), 231.
- Steel cores, solid, magnetic time-lag effects in. H. ASPDEN, (P), 279.
- (sheet), electrical, eddy-current anomaly in. H. ASPDEN, (P), 272.
- Strength, electric, of transformer oil. M. E. Z. EL-DINE and H. TROPPER, (P), 35.
- Superconducting sphere, magnetic energy and electron inertia in. E. G. CULLWICK, (P), 441.
- Surface loss in laminated pole-face. (D), 230.
- Switching transients, suppression of, by shunt RC circuit. (D), 233.
- Synthesizer, continuous delay-line, as system analogue. J. H. WESTCOTT, (P), 357.
- , discontinuous delay-line, properties of feedback-system analogue based on. R. M. F. HOUTAPPEL, (P), 367.
- System analogue, continuous delay-line synthesizer as. J. H. WESTCOTT, (P), 357.

T

- Tables and properties of Laguerre functions. J. W. HEAD and W. P. WILSON, (P), 428.
- Tape helices, wavelength measurements on. C. P. ALLEN and G. M. CLARKE, (P), 171.
- Teeth of slotted armature, reluctance of. S. NEVILLE, (P), 338.
- Terminology and notation in information theory, I. J. GOOD, (P), 200.
- Third-harmonic and zero-sequence fields, theory of. G. H. RAWCLIFFE and B. C. McDERMOTT, (P), 212.
- Three-phase systems, simultaneous faults in. (D), 233.
- Time-constant of self-saturating (auto-excited) transducers. U. KRABBE, (P), 71.
- lag effects in solid steel cores. H. ASPDEN, (P), 279.
- Tones (interfering) and noise, discrimination of synchronized oscillator against. D. G. TUCKER and G. G. JAMIESON, (P), 129.
- Torque components in the Schrage motor operating at synchronous speed. I. THOMAS, (P), 11.
- limitation, integral control with. J. C. WEST and M. J. SOMERVILLE, (P), 407.
- operated wattmeter. (See Wattmeter.)
- Transducers, a.c.-controlled. A. G. MILNES and T. S. LAW, (P), 81.
- , self-saturating (auto-excited), residual time-constant of. U. KRABBE, (P), 71.
- Transfer function of unbalanced RC networks, properties of. I. CEDERBAUM, (P), 400.
- Transformer oil, electric strength of. M. E. Z. EL-DINE and H. TROPPER, (P), 35.
- Transmitting-valve filament structures, mutual heating in. W. J. POHL, (P), 224.
- TROPPER, H., and EL-DINE, M. E. Z. (See EL-DINE.)
- TUCKER, D. G., and JAMIESON, G. G. Discrimination of synchronized oscillator against interfering tones and noise. (P), 129.

V

- VAUGHAN, J. R. M. Millimetre-wave magnetron. (P), 95.

W

- WAIT, J. R., and KATES, J. Radiation patterns of circumferential slots on moderately large conducting cylinders. (P), 289.
- Wattmeter, torque-operated, for microwave power. R. A. BAILEY, (P), 59.
- Wave, electromagnetic, diffraction of, by aperture in plane screen. R. F. MILLAR, (P), 177.
- Waveguide junction, axially unsymmetrical, calculation of equivalent circuit of. R. E. COLLIN and J. BROWN, (P), 121.
- Waveguides, anisotropic, microwave propagation in. A. E. KARBOWIAK, (P), 139.
- , ferromagnetic, attenuation and permeability of, between 9 000 and 9 675 Mc/s. J. ALLISON and F. A. BENSON, (P), 205.
- of arbitrary cross-sections, junction admittance between. E. D. FARMER, (P), 145.
- Wavelength measurements on tape helices, interpretation of. C. P. ALLEN and G. M. CLARKE, (P), 171.
- Waves, impulse voltage, peak-voltage measurements of. A. AKED, (P), 186.
- , radio, reflected from ionosphere at oblique incidence, polarization of. W. C. BAIN and C. B.-I. GLASS, (P), 447.
- WERNER, J. F., BIRCH, J., FRITH, A. G., FERGUSON, A. C. L., and MILES, R. H. A. (See BIRCH.)
- WEST, J. C., and SOMERVILLE, M. J. Integral control with torque limitation. (P), 407.
- WESTCOTT, J. H. Continuous delay-line synthesizer as system analogue. (P), 357.
- WILLSHAW, W. E., and ROBERTSHAW, R. G. (See ROBERTSHAW.)
- WILSON, W. P., and HEAD, J. W. (See HEAD.)
- WOOTTON, D. J., and PEARCE, A. F. Reflex klystron oscillator for 8-9 mm band. (P), 104.

Z

- ZEIN EL-DINE. (See EL-DINE.)
- Zero-sequence fields, theory of. RAWCLIFFE G. H. and McDERMOTT, B. C. (P), 212.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

PART C—MONOGRAPHS, SEPTEMBER 1956

CONTENTS

	PAGE
Discussion on 'The Suppression of Switching Transients by a Shunt RC Circuit'.....	233
Discussion on 'Matrix Methods for the Evaluation of Simultaneous Faults in Three-Phase Systems'.....	233
Relative Dynamic Stability of Large Synchronous Generators..... H. K. MESSERLE, M.Eng.Sc., B.E.E. (No. 159)	234
An Experimental Study of Intensity Spectra after Half-Wave Rectification of Signals in Noise.....	
G. E. FELLOWS, Ph.D., M.S., B.A., and D. MIDDLETON, Ph.D., M.A., A.B. (No. 160)	243
On the Inductance of Iron-Cored Coils..... P. HAMMOND, M.A. (No. 161)	249
An Approximate Method for Finding the 'Best Linear Servo Mechanism'..... H. H. ROSENBROCK, B.Sc.(Eng.), Ph.D. (No. 162)	260
Analysis of Linear n -Port Networks..... I. CEDERBAUM (No. 163)	267
The Eddy-Current Anomaly in Electrical Sheet Steel..... H. ASPDEN, Ph.D., B.Sc., Wh.Sc. (No. 164)	272
Magnetic Time-Lag Effects in Solid Steel Cores..... H. ASPDEN, Ph.D., B.Sc., Wh.Sc. (No. 165)	279
Note on the Fourier Coefficients for Chebyshev Patterns..... Dr. H. E. SALZER (No. 166)	286
Radiation Patterns of Circumferential Slots on Moderately Large Conducting Cylinders.....	
J. R. WAIT, M.A.Sc., Ph.D., and J. KATES, M.A., Ph.D. (No. 167)	289
Some Properties of Magnetrons using Spatial-Harmonic Operation..... R. G. ROBERTSHAW and W. E. WILLSHAW, M.B.E., M.Sc.Tech. (No. 168)	297
The Determination of the True Side-Lobe Level of Long Broadside Arrays from Radiation-Pattern Measurements made in the Fresnel Region.....	
R. H. T. BATES, B.Sc.(Eng.), and J. ELLIOTT, M.Sc. (No. 169)	307
Possible Errors of a Particular Wide-Aperture Direction-Finder..... W. C. BAIN, M.A., B.Sc., Ph.D. (No. 170)	313
An Audio-Frequency Dynamometer Wattmeter..... A. H. M. ARNOLD, Ph.D., D.Eng., and J. J. HILL, B.Sc. (No. 171)	325
Resultant Reactive Power of Overhead Lines..... A. F. COVENTRY, B.Sc. (No. 172)	334
Reluctance of the Teeth of a Slotted Armature..... S. NEVILLE, B.Sc.(Eng.) (No. 173)	338
The Use of the Blackburn A.C. Network Analyser in the Analysis of Power System Faults.....	
J. H. BANKS, M.Sc., and K. C. PARTON, B.Sc. (No. 174)	342
H.F. Bearing Variations on an Adcock Direction-Finder..... E. N. BRAMLEY, M.Sc., Ph.D. (No. 175)	350
The Continuous Delay-Line Synthesizer as a System Analogue..... J. H. WESTCOTT, B.Sc.(Eng.), Ph.D. (No. 176)	357
Properties of a Feedback-System Analogue based on a Discontinuous Delay-Line Synthesizer..... R. M. F. HOUTAPPEL, Dr.Ir. (No. 177)	367
The Application of the Frequency-Response Method to Electrical Machines..... S. K. SEN, B.E., and B. ADKINS, M.A. (No. 178)	378
Microwave Measurements with a Lossy Variable Termination..... H. M. ALTSCHULER, M.E.E., and Prof. A. A. OLINER, Ph.D. (No. 179)	392
Some Properties of the Transfer Function of Unbalanced RC Networks..... I. CEDERBAUM (No. 180)	400
Integral Control with Torque Limitation..... J. C. WEST, Ph.D., and M. J. SOMERVILLE, B.Sc. (No. 181)	407
The Use of Ethylene Diamine Tartrate for Piezo-Electric Filter Elements.....	
J. BIRCH, A. G. FRITH, A. C. L. FERGUSON, R. H. A. MILES, M.A., B.Sc., and J. F. WERNER, B.Sc. (No. 182)	420
Laguerre Functions: Tables and Properties..... J. W. HEAD, M.A., and W. PROCTOR WILSON, C.B.E., B.Sc.(Eng.) (No. 183)	428
Magnetic Energy and Electron Inertia in a Superconducting Sphere..... Prof. E. G. CULLWICK, O.B.E., M.A., D.Sc. (No. 184)	441
The Polarization of Very Long Radio Waves reflected from the Ionosphere at Oblique Incidence.....	
W. C. BAIN, M.A., B.Sc., Ph.D., and C. B.-I. GLASS, B.Sc., Ph.D. (No. 185)	447
A Study of the Field Distribution at an Axial Focus of a Square Microwave Lens..... P. A. MATTHEWS, B.Sc., and Prof. A. L. CULLEN, Ph.D. (No. 186)	449

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which the Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) which are out of print and from which reprints are not available. The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part. Example.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

The Benevolent Fund



Have YOU yet responded to the appeal for contributions to the

HOMES FUND

The Court of Governors hope that every member will contribute to this worthy object

Contributions may be sent by post to

THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF
ELECTRICAL ENGINEERS, SAVOY PLACE, LONDON, W.C.2

or may be handed to one of the Local Hon. Treasurers of the Fund



Local Hon. Treasurers of the Fund:

EAST MIDLAND CENTRE R. C. Woods
IRISH BRANCH A. Harkin, M.E.
MERSEY AND NORTH WALES CENTRE . . . D. A. Picken
NORTH-EASTERN CENTRE J. F. Skipsey, B.Sc.
NORTH MIDLAND CENTRE J. G. Craven
SHEFFIELD SUB-CENTRE F. Seddon
NORTH-WESTERN CENTRE W. E. Swale
NORTH LANCASHIRE SUB-CENTRE
G. K. Alston, B.Sc.(Eng.)

NORTHERN IRELAND CENTRE G. H. Moir, J.P.
SCOTTISH CENTRE R. H. Dean, B.Sc.Tech.
NORTH SCOTLAND SUB-CENTRE P. Philip
SOUTH MIDLAND CENTRE W. E. Clark
RUGBY SUB-CENTRE H. Orchard
SOUTHERN CENTRE G. D. Arden
WESTERN CENTRE (BRISTOL) A. H. McQueen
WESTERN CENTRE (CARDIFF) D. J. Thomas
WEST WALES (SWANSEA) SUB-CENTRE . . . O. J. Mayo
SOUTH-WESTERN SUB-CENTRE W. E. Johnson

THE BENEVOLENT FUND

Published by The Institution, Savoy Place, London, W.C.2.

Telephone: Temple Bar 7676.

Printed by Unwin Brothers Limited, Woking and London.

Telegrams: 'Voltampere, Phone, London.'

DATE DUE

JUL 15 1982

**PERIODICALS MUST BE RETURNED
TO PERIODICALS DESK ONLY**

DEMCO 38227